



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 116077

TO: Bao-Qun Li  
Location: REM-3D24  
Art Unit: 1648  
Sunday, March 07, 2004

3C/18

Case Serial Number: 09/643458

From: Toby Port  
Location: Biotech-Chem Library  
Remsen 1A59  
Phone: 571-272-2523

toby.port@uspto.gov

### Search Notes

Dear Examiner Li,

Here are the results of your search.  
Please feel free to contact me if you have any questions.

Toby Port

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model  
on: March 7, 2004, 13:27:25 ; Search time 58.5494 Seconds  
(without alignments)  
178.554 Million cell updates/sec

File: US-09-643-458B-9  
Effect score: 193  
Sequence: 1 CKKGGAATALEINLEEDDDNEDVDEQAEQKTHVF 37

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Aligned: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: A\_Geneseq\_29Jan04.\*

- 1: Geneseqp1980s.\*
- 2: Geneseqp1990s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	167	86.5	189	2	AAW79541 Adenoviru
2	167	86.5	338	6	AAO23323 Human Hu5
3	167	86.5	338	6	ABP56575 Human ade
4	167	86.5	952	2	AAW79539 Adenoviru
5	167	86.5	952	2	AAW63118 Human ade
6	162	83.9	45	2	AAW79543 Adenoviru
7	66.5	34.5	344	3	AAAG16071 Arabidops
8	66.5	34.5	348	3	AAAG16070 Arabidops
9	66.5	34.5	354	3	AAAG16069 Arabidops
10	66	34.2	635	4	ABG22101 Novel hum
11	66	34.2	727	5	ABP97597 Novel hum
12	66	34.2	728	7	ADD48422 Human Pro
13	66	34.2	728	7	ADD48426 Human Pro
14	66	34.2	745	4	ABG22102 Novel hum
15	66	34.2	764	2	AAW79531 Human lun
16	66	34.2	764	3	AAAB4430 Human lun
17	66	34.2	764	4	AAE13771 Human lun
18	66	34.2	764	7	AD66375 Human lun
19	66	34.2	764	7	ABE87629 Human lun
20	66	34.2	765	5	ABH57135 Mouse isc
21	66	34.2	904	4	ABG22103 Novel hum
22	66	34.2	1058	4	ABE71401 Drosophil
23	65.5	33.9	230	5	ABG93275 C. albica
24	65	33.7	712	7	AD61623 Rat Prote
25	65	33.7	712	7	ADD46090 Rat Prote

## ALIGNMENTS

RESULT 1  
AAW79541  
ID AAW79541 standard; protein; 189 AA.  
XX  
AC AAW79541;  
XX  
DT 17-OCT-2003 (revised)  
DT 11-JAN-1999 (first entry)  
XX  
DE Adenovirus serotype 5 loop 1 region.  
XX  
KW Adenovirus serotype 5; Ad5; hexon; coat protein; gene therapy; vector.  
XX  
OS Human adenovirus type 5.  
XX  
PN WO9840509-A1.  
XX  
PD 17-SEP-1998.  
XX  
PF 13-MAR-1998; 98WO-US005033.  
XX  
PR 13-MAR-1997; 97US-00816346.  
XX  
(CORR ) CORNELL RES FOUND INC.  
(GENV-) GENVEC INC.  
XX  
Crystal RG, Falck-Pedersen E, Gall J, Kovessdi I, Wickham TJ;  
WPI; 1998-506738/43.  
N-PSDB; AAV61503.  
XX  
Chimeric adenovirus coat protein - useful in, e.g. vector for gene transfer to treat inherited genetic diseases.  
XX  
Claim 7; Page 75-76; 112pp; English.

This is the amino acid sequence of the loop 1 (11) region of the hexon protein of the adenovirus serotype 5 (Ad5) coat protein (see also AAW79539). It is encoded by a claimed DNA sequence (see AAV61503). The invention provides a chimeric adenoviral coat protein, particularly a chimeric adenovirus hexon protein, that has a decreased ability or inability to be recognised by a neutralising antibody directed against the corresponding wild-type adenovirus coat protein. The chimeric adenoviral coat protein has a non-native amino acid sequence, especially a comprising a deletion of an internal hexon protein sequence, preferably a hypervariable region or entire loop. DNA sequences (see AAV61503-23) encoding claimed Ad2 and Ad5 chimeric coat proteins (see AAW79540-61) are claimed. Also claimed are an adenovirus vector that comprises the

Ad57828 Rat Prote  
Ad57830 Rat Prote  
Ad564235 Human pro  
Aaw79542 Adenoviru  
Abg76380 Translati  
Abg29213 Novel hum  
Aaw79540 Adenoviru  
Abp56672 Human ade  
Aaw94764 Rat type  
Aaw94763 Mouse typ  
Abg76382 Rat type  
Abg76381 Rat type  
Aaw79538 Adenoviru  
Abg76372 Translati  
Abp33416 Human red  
Aar99631 Eucalyptu  
Abg06324 Novel hum  
Abg93099 S. cerevi  
Abg23062 Novel hum  
Aag16884 Arabidops

26 65 33.7 712 7 ADE57828  
27 65 33.7 712 7 ADE57830  
28 64.5 33.4 661 7 ADB64235  
29 64 33.2 51 2 AAW79542  
30 64 33.2 71 6 ABG76380  
31 64 33.2 74 4 ABG29213  
32 64 33.2 201 2 AAW79540  
33 64 33.2 345 6 ABP56672  
34 64 33.2 478 2 AAW94764  
35 64 33.2 478 2 AAW94763  
36 64 33.2 480 6 ABG76382  
37 64 33.2 480 6 ABG76381  
38 64 33.2 968 2 AAW79538  
39 63 32.6 71 6 ABG76372  
40 63 32.6 116 5 ABP33416  
41 63 32.6 119 2 AAR99631  
42 63 32.6 147 4 ABG06324  
43 63 32.6 147 5 ABG93099  
44 63 32.6 205 4 ABG23062  
45 63 32.6 374 3 AAG16884



CC and antimicrobial activities, and can be used in vaccines. The chimpanzee  
 CC C68 adenoviral capsid protein is useful for preparing a composition for  
 CC treating hyperproliferative conditions e.g., cancer or psoriasis and as a  
 CC vaccine against bacterial, fungal, viral or parasitic infection. The  
 CC present sequence represents a human adenovirus capsid hexon protein which  
 CC is given in comparison with the chimpanzee C68 adenovirus hexon protein  
 CC in the exemplification of the present invention

CC Sequence 338 AA;

Query Match 86.5%; Score 167; DB 6; Length 338;

Best Local Similarity 89.2%; Pred. No. 3e-13; Indels 0; Gaps 0;

Matches 33; Conservative 1; Mismatches 3;

1 CKGGAATALEINLEERDDDDNEDEVDEQAQQKTHVF 37

9 CEWDEAATALEINLEERDDDDNEDEVDEQAQQKTHVF 45

RESULT 4

AAW79539

ID AAW79539 standard; protein; 952 AA.

AC AAW79539;

CC AAW79539;

CT 17-OCT-2003 (revised)

CT 11-JAN-1999 (first entry)

CT Adenovirus serotype 5 hexon protein.

CT Adenovirus serotype 5; Ad5; hexon; coat protein; gene therapy; vector.

CT Human adenovirus type 5.

CT Key

CT Region

CT /label= l1 loop

CT /note= "Claim 7"

CT Region

CT /label= HVR1 region

CT /note= "Claim 7"

CT Region

CT /label= HVR2 region

CT /note= "Claim 7"

CT Region

CT /label= HVR3 region

CT /note= "Claim 7"

CT Region

CT /label= HVR4 region

CT /note= "Claim 7"

CT Region

CT /label= HVR5 region

CT /note= "Claim 7"

CT Region

CT /label= HVR6 region

CT /note= "Claim 7"

CT Region

CT /label= l2 loop

CT /note= "Claim 7"

CT Region

CT /label= HVR7 region

CT /note= "Claim 7"

CT Region

CT Misc-difference 951

CT /label= Gln, His, Thr

CT Misc-difference 952

CT /label= Gln, His, Thr

CT WO9840509-A1.

CT 17-SEP-1998.

CT 13-MAR-1998;

CT 98WO-US0005033.

CT

PR 13-MAR-1997; 97US-00816346.

XX (CORR ) CORNELL RES FOUND INC.

PA (GENV-) GENVEC INC.

XX Crystal RG, Falck-Pedersen E, Gall J, Kovesdi I, Wickham TJ;

XX WPI; 1998-506738/43.

XX N-PSDB; AAV61501.

XX Chimeric adenovirus coat protein - useful in, e.g. vector for gene

XX transfer to treat inherited genetic diseases.

XX Disclosure; Page 66-70; 112pp; English.

XX This is the amino acid sequence of the wild-type hexon protein of

XX adenovirus serotype 5 (Ad5). The invention provides a chimeric adenoviral

XX coat protein, particularly a chimeric adenovirus hexon protein, that has

XX a decreased ability or inability to be recognised by a neutralising

XX antibody directed against the corresponding wild-type adenovirus coat

XX protein. The chimeric adenoviral coat protein has a non-native amino acid

XX sequence, especially comprising a deletion of one or more regions of the

XX hexon protein, particularly a hypervariable region (HVR) of loop 1 or 2

XX of the hexon protein. DNA sequences (see AAV61502-23) encoding claimed

XX Ad2 and Ad5 chimeric coat proteins (see AAV79540-61) are claimed. Also

XX claimed are an adenovirus vector that comprises the chimeric adenovirus

XX coat protein, a method of genetically modifying a cell by contacting with

XX the vector, and a host cell that comprises the chimeric adenovirus coat

XX protein. The vector can be used for gene transfer, for the treatment of

XX inherited diseases (e.g. by carrying the cystic fibrosis (CF)

XX transmembrane conductance regulator cDNA to treat CF). It can also be

XX used to render certain cells susceptible to the killing action of certain

XX drugs, or to study the effects of expression of specific genes in a given

XX cell or tissue in vitro or in vivo. (Updated on 17-OCT-2003 to

XX standardise OS field)

XX Sequence 952 AA;

XX Query Match 86.5%; Score 167; DB 2; Length 952;

XX Best Local Similarity 89.2%; Pred. No. 9.8e-13;

XX Matches 33; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

XX

QY 1 CKGGAATALEINLEERDDDDNEDEVDEQAQQKTHVF 37

Db 133 CEWDEAATALEINLEERDDDDNEDEVDEQAQQKTHVF 169

RESULT 5

AAW63118

ID AAW63118 standard; protein; 952 AA.

XX AAW63118;

XX 17-OCT-2003 (revised)

XX 24-NOV-1998 (first entry)

XX Human adenovirus 5 hexon.

XX Adenovirus 5; Ad5; hexon; vector; gene therapy; vaccine.

XX Human adenovirus type 5.

XX Key

XX Location/Qualifiers

XX Misc-difference 88

XX /note= "encoded by GGG"

XX Region

XX /label= L1

XX /note= "loop 1"

XX Region

XX /label= L2

XX /note= "loop 2"

XX Region

XX /label= L4

FT Misc-difference 602 /note= "loop 4"  
 FT /note= "encoded by ACG"

PN WO9832842-A1.

XX 30-JUL-1998.

XX 22-JAN-1998; 98WO-US001113.

XX 24-JAN-1997; 97US-00788674.

XX (GENE-) GENETIC THERAPY INC.

XX Roy S;

XX WPI; 1998-427938/36.

XX N-PSDB; AAV42656.

XX Modified adenoviral vaccine - comprises adenovirus with a hexon loop region of a second viral serotype.

XX Disclosure; Fig 1Aa-e; 67pp; English.

XX This is the amino acid sequence of the adenovirus 5 hexon. A modified adenovirus, in which the original adenovirus is of a first serotype within a first subgenus, and in which, in the modified version, at least a portion of at least one loop region of the hexon is removed and replaced with at least a portion of at least one loop region of the hexon of an adenovirus of a second serotype within a second subgenus, is new. Also claimed is a method of providing a therapeutic effect in a host comprising the administration of the adenovirus. The invention can be used as a vaccine against adenoviruses or as a gene therapy vector. The adenovirus 12 hexon amino acid sequence is also provided (see AAW63117). (Updated on 17-OCT-2003 to standardise OS field)

XX Sequence 952 AA;

Query Match 86.5%; Score 167; DB 2; Length 952;

Best Local Similarity 89.2%; Pred. No. 9.8e-13; Indels 0; Gaps 0; Matches 33; Conservative 1; Mismatches 3;

QY 1 CKGKAATALEINLEEDDDNEDEVDQAEQOKTHVF 37  
 Db 133 CEWDEAATALEINLEEDDDNEDEVDQAEQOKTHVF 169

RESULT 6

AAW79543

XX AAW79543 standard; protein; 45 AA.

XX AC AAW79543;

XX 17-OCT-2003 (revised)

XX 11-JAN-1999 (first entry)

XX Adenovirus serotype 5 hypervariable region HVRI.

XX Adenovirus serotype 5; Ad5; hexon; coat protein; gene therapy; vector.

XX Human adenovirus type 5.

XX WO9840509-A1.

XX 17-SEP-1998.

XX 13-MAR-1998; 98WO-US005033.

XX 13-MAR-1997; 97US-00816346.

XX (CORR ) CORNELL RES FOUND INC.

XX (GENV-) GENVEC INC.

XX

PI Crystal RG, Falck-Pedersen E, Gall J, Kovsdi I, Wickham TJ;

XX WPI; 1998-506738/43.

XX N-PSDB; AAV61505.

XX Chimeric adenovirus coat protein - useful in, e.g. vector for gene transfer to treat inherited genetic diseases.

XX Claim 7; Page 77; 112pp; English.

XX This is the amino acid sequence of hypervariable region HVRI of the hexon protein (see AAW79539) of adenovirus serotype 5 (Ad5). It is encoded by a claimed DNA sequence (see AAV61505). The invention provides a chimeric adenoviral coat protein, particularly a chimeric adenovirus hexon protein, that has a decreased ability or inability to be recognised by a neutralising antibody directed against the corresponding wild-type adenovirus coat protein. The chimeric adenoviral coat protein has a non-native amino acid sequence, especially comprising a deletion of an internal hexon protein sequence, preferably a hypervariable region or entire loop. DNA sequences (see AAV61502-23) encoding claimed Ad2 and Ad5 chimeric coat proteins (see AAW79540-61) are claimed. Also claimed are an adenovirus vector that comprises the chimeric adenovirus coat protein, a method of genetically modifying a cell by contacting it with the vector, and a host cell that comprises the chimeric adenovirus coat protein. The vector can be used for gene transfer, for the treatment of inherited diseases. It can also be used to render certain cells susceptible to the killing action of certain drugs, or to study the effects of expression of specific genes in a given cell or tissue in vitro or in vivo. (Updated on 17-OCT-2003 to standardise OS field)

XX Sequence 45 AA;

Query Match 83.9%; Score 162; DB 2; Length 45;

Best Local Similarity 100.0%; Pred. No. 1.3e-13;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AATALEINLEEDDDNEDEVDQAEQOKTHVF 37  
 Db 1 AATALEINLEEDDDNEDEVDQAEQOKTHVF 32

RESULT 7

AAAG16071

XX AAG16071 standard; protein; 344 AA.

XX AC AAG16071;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 16570.

XX Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

XX 05-MAR-1999; 99US-0123180P.

XX 09-MAR-1999; 99US-0123548P.

XX 23-MAR-1999; 99US-0125788P.

XX 25-MAR-1999; 99US-0126264P.

XX 29-MAR-1999; 99US-0126785P.

XX 01-APR-1999; 99US-0127462P.

XX 06-APR-1999; 99US-0128234P.

XX 08-APR-1999; 99US-0128714P.

XX 16-APR-1999; 99US-0129845P.



```
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 34.5%; Score 66.5; DB 3; Length 344;
Best Local Similarity 37.9%; Pred. No.2.9;
Matches 11; Conservative 11; Mismatches 6; Indels 1; Gaps 1;

QY 5 GAATALEINLEEDDDNDEVDDEQAEOQK 33
DB 144 GSVTG-DVVDNDNDDNDEDDDDSEDE 171

RESULT 8
ID AAG16070 standard; protein; 348 AA.
XX AC AAG16070;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 16569.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PD EP1033405-A2.
XX PF 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX PF 25-FEB-1999; 99US-0121825P.
XX PF 05-MAR-1999; 99US-0123180P.
XX PF 09-MAR-1999; 99US-0123548P.
XX PF 23-MAR-1999; 99US-0125788P.
XX PF 25-MAR-1999; 99US-0126264P.
XX PF 29-MAR-1999; 99US-0126785P.
XX PF 01-APR-1999; 99US-0127462P.
XX PF 06-APR-1999; 99US-0128234P.
XX PF 08-APR-1999; 99US-0128714P.
XX PF 16-APR-1999; 99US-0128845P.
XX PF 19-APR-1999; 99US-0130077P.
XX PF 21-APR-1999; 99US-0130449P.
XX PF 23-APR-1999; 99US-0130510P.
XX PF 23-APR-1999; 99US-0130891P.
XX PF 28-APR-1999; 99US-0131449P.
XX PF 30-APR-1999; 99US-0132048P.
XX PF 30-APR-1999; 99US-0132407P.

PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137503P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139466P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 06-JUL-1999; 99US-0142053P.
PR 08-JUL-1999; 99US-0142390P.
PR 09-JUL-1999; 99US-0142803P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143542P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 19-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
```





PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135623P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137503P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 24-JUN-1999; 99US-0140696P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 02-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 06-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145219P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 08-AUG-1999; 99US-0147433P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 09-AUG-1999; 99US-0148171P.  
PR 10-AUG-1999; 99US-0148319P.  
PR 11-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 15-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 23-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-015659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 08-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.





The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 728 AA;

Query Match 34.28; Score 66; DB 7; Length 728;  
Best Local Similarity 40.09; Fred.No. 7.8;  
Matches 12; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

4 KGAATALEINLEEDDDNDEVDQAEQQK 33  
| : | : | : | : | : | :  
628 KSSRTTILQSKSEDEDEDEDDEEE 657

RESULT 14  
ABG22102  
ID ABG22102 standard; protein; 745 AA.  
XC  
XC CC ABG22102;  
XC  
XT 18-FEB-2002 (first entry)  
XE Novel human diagnostic protein #22093.  
XW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XW food supplement; medical imaging; diagnostic; genetic disorder.  
XS Homo sapiens.  
XT WO200175067-A2.  
XD 11-OCT-2001.  
XF 30-MAR-2001; 2001WO-US008631.  
XR 31-MAR-2000; 2000US-00540217.  
XR 23-AUG-2000; 2000US-00649167.  
XA (HYSE-) HYSEQ INC.  
XI Drmanac RT, Liu C, Tang YT;  
XR WPI; 2001-639362/73.  
XR N-PSTB; AAS86289.  
  
New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.  
  
Claim 20; SEQ ID NO 52461; 103pp; English.  
  
The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG0377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at

[illegible]

Search completed: March 7, 2004, 13:35:08  
Job time : 59.5494 secs

---



C:Superfamily: nucleolin; ribonucleoprotein repeat homology  
 C:Keywords: DNA binding; nucleus; phosphoprotein  
 F:234-299/Domain: ribonucleoprotein repeat homology <RRM1>  
 F:326-389/Domain: ribonucleoprotein repeat homology <RRM2>  
 F:416-478/Domain: ribonucleoprotein repeat homology <RRM3>  
 F:504-568/Domain: ribonucleoprotein repeat homology <RRM4>

Query Match 43.0%; Score 83; DB 2; Length 651;  
 Best Local Similarity 54.8%; Pred. No. 0.15;  
 Matches 17; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 2 KKGGAATALEINLEEDDDNEDEVEDEQAEQQ 32  
 DB 175 KGGKTAQAABEDDEDDDEDDDEDEDEEQ 205

RESULT 3  
 T22846  
 hypothetical protein F57C7.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
 C:Accession: T22846  
 R:White, S.  
 submitted to the EMBL Data Library, February 1996  
 A:Reference number: Z19625  
 A:Accession: T22846  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-288 <WIL>  
 A:Cross-references: EMBL:Z69646; PIDN:CAA93474.1; GSPDB:GN00028; CESP:F57C7.3  
 A:Experimental source: clone F57C7  
 C:Genetics:  
 A:Gene: CESP:F57C7.3  
 A:Map position: X  
 A:Introns: 51/3; 87/1; 120/3; 191/2; 231/1

Query Match 35.2%; Score 68; DB 2; Length 288;  
 Best Local Similarity 40.6%; Pred. No. 2.8;  
 Matches 13; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

QY 2 KKGGAATALEINLEEDDDNEDEVEDEQAEQQ 33  
 DB 129 KKPANDKEIKVEDEDDDEDEDEDEDEDEZ 150

RESULT 4  
 T46608  
 zinc finger protein Png-1 - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 18-Feb-2000 #sequence\_revision 18-Feb-2000 #text\_change 18-Feb-2000  
 C:Accession: T46608  
 R:Weiner, J.; Chun, J.  
 J. Comp. Neurol. 381, 130-142, 1997  
 A:Title: Png-1, a nervous system-specific zinc finger gene, identifies regions containin  
 A:Reference number: Z23102; MUID:97276971; PMID:9130664  
 A:Accession: T46608  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1188 <WEI>  
 A:Cross-references: EMBL:U86338; NID:g1835754; PID:g1835755; PIDN:AAC53157.1  
 A:Experimental source: strain BALB/c  
 C:Genetics:  
 A:Gene: Png-1

Query Match 35.2%; Score 68; DB 2; Length 1188;  
 Best Local Similarity 44.0%; Pred. No. 1.1;  
 Matches 11; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 11 EINLEEDDDNEDEVEDEQAEQQKTH 35  
 DB 151 DVVEEDDDDEEEEEEEEEEEDH 175

RESULT 5  
 S18193  
 autoantigen NOR-90 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 24-Sep-1999  
 C:Accession: S18193  
 R:Chan, E.K.L.; Imai, H.; Hamel, J.C.; Tan, E.M.  
 J. Exp. Med. 174, 1239-1244, 1991  
 A:Title: Human autoantibody to RNA polymerase I transcription factor HUBF. Molecular ide  
 A:Reference number: S18193; MUID:92044316; PMID:1940801  
 A:Accession: S18193  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-727 <CHA>  
 A:Cross-references: EMBL:X56687; NID:g28970; PIDN:CAA40016.1; PID:g28971  
 C:Superfamily: unassigned HMG box proteins; HMG box homology  
 F:109-184/Domain: HMG box homology <HMG1>  
 F:193-231/Domain: HMG box homology #status atypical <HMG2>  
 F:257-329/Domain: HMG box homology <HMG3>  
 F:367-442/Domain: HMG box homology <HMG4>  
 F:445-516/Domain: HMG box homology <HMG5>  
 F:527-601/Domain: HMG box homology <HMG6>

Query Match 34.2%; Score 66; DB 2; Length 727;  
 Best Local Similarity 40.0%; Pred. No. 12;  
 Matches 12; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 4 KGAATALEINLEEDDDNEDEVEDEQAEQQ 33  
 DB 628 KSRRTTLOSKSESEDEDEDEDEDEDEE 657

RESULT 6  
 S09318  
 transcription factor UBF, nucleolar - human  
 C:Species: Homo sapiens (man)  
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 24-Sep-1999  
 C:Accession: S09318  
 R:Jantzen, H.M.; Admon, A.; Bell, S.P.; Tjian, R.  
 Nature 344, 830-836, 1990  
 A:Title: Nucleolar transcription factor HUBF contains a DNA-binding motif with homology  
 A:Reference number: S09318; MUID:90231434; PMID:2330041  
 A:Accession: S09318  
 A:Molecule type: mRNA  
 A:Residues: 1-764 <JAN>  
 A:Cross-references: EMBL:X53461; NID:g37573; PIDN:CAA37548.1; PID:g37574  
 A:Note: part of this sequence was confirmed by amino acid sequencing  
 C:Superfamily: unassigned HMG box proteins; HMG box homology  
 C:Keywords: DNA binding; transcription regulation  
 F:109-184/Domain: HMG box homology <HMG1>  
 F:193-268/Domain: HMG box homology <HMG2>  
 F:294-366/Domain: HMG box homology <HMG3>  
 F:404-479/Domain: HMG box homology <HMG4>  
 F:482-553/Domain: HMG box homology <HMG5>  
 F:564-638/Domain: HMG box homology <HMG6>

Query Match 34.2%; Score 66; DB 2; Length 764;  
 Best Local Similarity 40.0%; Pred. No. 12;  
 Matches 12; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 4 KGAATALEINLEEDDDNEDEVEDEQAEQQ 33  
 DB 665 KSRRTTLOSKSESEDEDEDEDEDEDEE 694

RESULT 7  
 S22314  
 transcription factor UBF - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 24-Sep-1999  
 C:Accession: S22314  
 R:Hisatake, K.; Nishimura, T.; Maeda, Y.; Hanada, K.; Song, C.Z.; Muramatsu, M.

uclic Acids Res. 19, 4631-4637, 1991  
A:Title: Cloning and structural analysis of cDNA and the gene for mouse transcription factor  
A:Reference number: S22314; MUID:91367658; PMID:1891354  
A:Accession: S22314  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-765 <EHL>  
A:Cross-references: EMBL:X50831; NID:G55115; PIDN:CAA43222.1; PID:G55116  
A:Note: The authors translated the codon ACG for residue 242 as Arg, GAG for residue 311  
A:Superfamily: unassigned HMG box proteins; HMG box homology  
A:Keywords: DNA binding; transcription regulation  
A:109-184/Domain: HMG box homology <HMG1>  
A:193-268/Domain: HMG box homology <HMG2>  
A:294-366/Domain: HMG box homology <HMG3>  
A:404-479/Domain: HMG box homology <HMG4>  
A:482-553/Domain: HMG box homology <HMG5>  
A:564-638/Domain: HMG box homology <HMG6>  
Query Match 34.2%; Score 66; DB 2; Length 765;  
Best Local Similarity 36.7%; Pred. No. 12;  
Matches 11; Conservative 10; Mismatches 9; Indels 0; Gaps 0;  
Y 4 KGAATALEINLEERDDNEDEVEDEQAQQK 33  
b 665 KSRRTLQKSESEDDDEEDDEDEEE 694  
RESULT 8  
84587  
ypothetical protein At2g20280 [imported] - Arabidopsis thaliana  
A:Species: Arabidopsis thaliana (mouse-ear cress)  
A:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
A:Accession: C84587  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-371 <STO>  
A:Cross-references: GB:AE002093; NID:G4512707; PIDN:AAD21760.1; GSPDB:GN00139  
A:Gene: At2g20280  
A:Map position: 2  
Query Match 33.9%; Score 65.5; DB 2; Length 371;  
Best Local Similarity 39.0%; Pred. No. 6.8;  
Matches 16; Conservative 8; Mismatches 10; Indels 7; Gaps 2;  
Y 3 GKGATALEINLEERDDNE-----DEVEDEQAQQKTHV 36  
b 320 GBAEQSKVEEEDDDDDDDDLMDLDELEASLSKTSI 360  
RESULT 9  
H0148  
nucleolin - rat  
A:Species: Rattus norvegicus (Norway rat)  
A:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 23-Jul-1999  
A:Accession: JH0148; A24088; I63130  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-187-196, 1990  
A:Gene: 88, 187-196, 1990  
A:Title: Nucleolin gene organization in rodents: highly conserved sequences within three  
A:Reference number: JH0148; MUID:90269607; PMID:2347493  
A:Accession: JH0148  
A:Molecule type: DNA  
A:Residues: 1-712 <BOU>  
A:Note: The authors translated the initiation codon GTG for residue 1 as Val  
A:Lischwe, M.A.; Cook, R.G.; Ahn, Y.S.; Yeoman, L.C.; Busch, H.  
A:Biochemistry 24, 6025-6028, 1985

A:Title: Clustering of glycine and Ng,Ng-dimethylarginine in nucleolar protein C23.  
A:Reference number: A24088; MUID:86104094; PMID:4084504  
A:Accession: A24088  
A:Molecule type: protein  
A:Residues: 651-703 <LIS>  
A:R:Bourbon, H.  
Gene 68, 73-84, 1988  
A:Title: Sequence and structure of the nucleolin promoter in rodents: Characterization of  
A:Reference number: I48118; MUID:89121496; PMID:2506027  
A:Accession: I63130  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 'MV', 2-44 <RES>  
A:Cross-references: GB:M22090; NID:G205793; PIDN:AAA41733.1; PID:G205794  
A:Comment: This protein is the major nucleolar-specific protein in eukaryotic exponential  
A:Genetics:  
A:Gene: nuc  
A:Start codon: GTG  
A:Introns: 5/3; 44/3; 211/1; 274/1; 303/1; 350/2; 392/1; 433/2; 485/1; 526/2; 571/1; 613/  
A:Superfamily: nucleolin; ribonucleoprotein repeat homology  
A:Keywords: DNA binding; nucleus  
A:311-376/Domain: ribonucleoprotein repeat homology <RRM1>  
A:397-459/Domain: ribonucleoprotein repeat homology <RRM2>  
A:489-552/Domain: ribonucleoprotein repeat homology <RRM3>  
A:575-639/Domain: ribonucleoprotein repeat homology <RRM4>  
Query Match 33.7%; Score 65; DB 2; Length 712;  
Best Local Similarity 40.6%; Pred. No. 15;  
Matches 13; Conservative 8; Mismatches 11; Indels 0; Gaps 0;  
QY 2 KKGATALEINLEERDDNEDEVEDEQAQQK 33  
Db 178 KAAPAASEDEDEDDDDDDDEDEEE 209  
RESULT 10  
T46637  
transcription factor 1, neural - rat  
A:Alternate names: neural zinc finger factor-1  
A:Species: Rattus norvegicus (Norway rat)  
A:Date: 18-Feb-2000 #sequence\_revision 18-Feb-2000 #text\_change 11-May-2000  
A:Accession: T46637  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1187 <JIA>  
A:Cross-references: EMBL:U48809; NID:G1511631; PID:G1511632; PIDN:AA052728.1  
A:Genetics:  
A:Gene: NZF-1  
A:Function:  
A:Description: binds specifically to a cis-regulatory element of the beta-retinoic acid ;  
e nervous system and in the pituitary gland  
A:Keywords: DNA binding; transcription factor  
Query Match 33.7%; Score 65; DB 2; Length 1187;  
Best Local Similarity 40.0%; Pred. No. 24;  
Matches 10; Conservative 8; Mismatches 7; Indels 0; Gaps 0;  
QY 11 EINLEEDDDNEDEVEDEQAQQKTH 35  
Db 150 DVEEDDDDEDEEEEDDEEDH 174  
RESULT 11  
JC4295  
heat-shock protein - slime mold (Dictyostelium discoideum)  
A:Species: Dictyostelium discoideum  
A:Date: 14-Nov-1995 #sequence\_revision 08-Feb-1996 #text\_change 29-Oct-1999  
A:Accession: JC4295





submitted to the EMBL Data Library, February 1998  
;Description: The sequence of C. elegans cosmid K07H8.  
;Reference number: Z21264  
;Accession: T33022  
;Status: preliminary; translated from GB/EMBL/DBJ  
;Molecule type: DNA  
;Residues: 1-798 <FUL>  
;Cross-references: EMBL:AF047659; PIDN:AAC04430.1; GSPDB:GN00022; CESP:K07H8.10  
;Experimental source: strain Bristol N2; clone K07H8  
;Genetics:  
;Gene: CESP:K07H8.10  
;Map position: 4  
;Intons: 205/1; 308/3; 630/1; 773/3  
  
Query Match 33.2%; Score 64; DB 2; Length 798;  
Best Local Similarity 40.7%; Pred. No. 21;  
Matches 11; Conservative 10; Mismatches 6; Indels 0; Gaps 0;  
  
y 7 ATALEINLEEDDDNEDEVQAEQOK 33  
b 421 AQAIEDSDDEDEDEDEDEDEEE 447  
  
earch completed: March 7, 2004, 13:38:24  
ob time : 15.2308 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

in on: March 7, 2004, 13:28:01 ; Search time 9.35165 Seconds  
(without alignments)  
206.017 Million cell updates/sec

itle: US-09-643-458B-9

erfect score: 193

quence: 1 CKGKGATALEINLEEDDDNEDEVDEAQEQKTHVF 37

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

sarched: 141681 seqs, 52070155 residues

otal number of hits satisfying chosen parameters: 141681

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

ssult No.	Score	Query Match	Length	DB	ID	Description
1	167	86.5	951	1	HEX_ADE05	P04133 human adeno
2	83	43.0	650	1	NUCL_XENLA	P20397 xenopus lae
3	68	35.2	288	1	SDC CAEEL	P50605 caenorhabdi
4	66	34.2	764	1	UBFI_HUMAN	P17480 homo sapien
5	66	34.2	765	1	UBFI_MOUSE	P25976 mus musculu
6	65.5	33.9	213	1	EF1B_CANAL	P78590 candida alb
7	65	33.7	712	1	NUCL_RAT	P13383 rattus norv
8	64.5	33.4	281	1	HG32_DICDI	P54658 dictyosteli
9	64.5	33.4	661	1	Y197_HUMAN	Q96P20 homo sapien
10	64	33.2	405	1	DEMA_MOUSE	Q9WV69 mus musculu
11	64	33.2	478	1	AMP2_RAT	P38062 rattus norv
12	64	33.2	706	1	NUCL_MOUSE	P09405 mus musculu
13	64	33.2	967	1	HEX_ADE02	P03277 human adeno
14	63	32.6	147	1	UCRH_YEAST	P00127 saccharomyc
15	63	32.6	239	1	CENB_SHEEP	P49451 ovis aries
16	63	32.6	359	1	FLI_EUCGL	O64953 eucalyptus
17	63	32.6	478	1	AMP2_MOUSE	O08663 mus musculu
18	63	32.6	797	1	VG48_HSV8A	Q01033 herpesvirus
19	63	32.6	1332	1	SPT7_YEAST	P35177 saccharomyc
20	62.5	32.4	694	1	NUCL_CHICK	P15771 gallus gall
21	62	32.1	199	1	HMGA_HUMAN	O15347 homo sapien
22	62	32.1	333	1	KC2B_NEUCR	Q8TG12 neurospora
23	62	32.1	992	1	VP41_YEAST	P38959 saccharomyc
24	61.5	31.9	630	1	YCP2_CENVI	P31569 oenothera v
25	61.5	31.9	721	1	YCP2_CENPI	P31568 oenothera p
26	61.5	31.9	764	1	UBFI_RAT	P25977 rattus norv
27	61	31.6	180	1	HMG1_CRIGR	P07156 cricetus
28	61	31.6	214	1	HMG1_MOUSE	P07155 mus musculu
29	61	31.6	279	1	ASF1_YEAST	P32447 saccharomyc
30	61	31.6	300	1	NXK1_BISBI	O46383 bison bison
31	61	31.6	465	1	HEX_ADE06	Q04966 human adeno
32	61	31.6	706	1	NUCL_HUMAN	P19338 homo sapien
33	61	31.6	1121	1	MYT1_HUMAN	Q01538 homo sapien

34 61 31.6 1790 1 US01\_YEAST P25386 saccharomyc  
35 60.5 31.3 382 1 SX11\_XENLA Q91731 xenopus lae  
36 60.5 31.3 1719 1 PRD2\_HUMAN Q13029 homo sapien  
37 60.5 31.3 1985 1 CCAF\_MOUSE Q9JIS7 mus musculu  
38 60 31.1 411 1 MP62\_LYTP1 P91753 lytechinus  
39 60 31.1 412 1 FK94\_SPOPR Q26486 spodoptera  
40 60 31.1 467 1 HEX\_ADE01 Q04965 human adeno  
41 60 31.1 489 1 ENGA\_PRESM Q885V6 pseudomonas  
42 60 31.1 582 1 PESB\_BRARE P79741 brachydanio  
43 60 31.1 623 1 ARE1\_SACBA Q87613 saccharomyc  
44 60 31.1 702 1 KF3A\_HUMAN Q9Y496 homo sapien  
45 60 31.1 758 1 CHLD\_TOBAC O24133 nicotiana t

#### ALIGNMENTS

RESULT 1  
HEX\_ADE05  
ID HEX\_ADE05 STANDARD; PRT; 951 AA.  
AC P04133;  
DT 01-NOV-1986 (Rel. 03, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hexon protein (late protein 2).  
GN PII.  
OS Human adenovirus type 5.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
OX NCBI\_TaxID=28285;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84212465; PubMed=6202684;  
RA Kinloch R., Mackay N., Mautner V.;  
RT "Adenovirus hexon. Sequence comparison of subgroup C serotypes 2 and 5."  
RL J. Biol. Chem. 259:6431-6436(1984).  
RN [2]  
RP COMPLETE GENOME.  
RX MEDLINE=92087470; PubMed=1727603;  
RA Chroboczek J., Bieber F., Jacrot B.;  
RT "The sequence of the genome of adenovirus type 5 and its comparison with the genome of adenovirus type 2."  
RL Virology 186:280-285(1992).  
CC -1- FUNCTION: This protein is one of the structural proteins in the viral coat and is synthesized during late infection.  
CC -1- SUBUNIT: Homotrimer (By similarity).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; J01966; AAD15293.1; --  
CC EMBL; M73260; --; NOT ANNOTATED\_CDS.  
CC EMBL; X02997; CAA26753.1; --  
CC PIR; A03849; HYAD5.  
CC PDB; 1P30; 1I-NOV-03.  
CC InterPro; IPR000736; Adeno\_hexon.  
CC Pfam; PF01065; Adeno\_hexon; 1.  
CC Pfam; PF03678; Adeno\_hexon; 1.  
CC ProDom; PD002815; Adeno\_hexon; 1.  
KW Coat protein; Hexon protein; Late protein; 3D-structure.  
FT INIT MET 0 0 BY SIMILARITY.  
FT DOMAIN 133 158 ASP/GLU-RICH (ACIDIC)  
SQ SEQUENCE 951 AA; 107875 MW; EF0F9A24961B117F CRC64;  
Query Match 86.5%; Score 167; DB 1; Length 951;  
Best Local Similarity 89.2%; Pred. No. 1.9e-10;  
Matches 33; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

2y 1 CKKGATALEINLEEDDDNEDEVDQAEQKTHVF 37  
 Db 132 CEWDRAATALEINLEEDDDNEDEVDQAEQKTHVF 168

## RESULT 2

NUCL\_XENLA STANDARD; PRT; 650 AA.  
 AC P20397;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Nucleolin (Protein C23).  
 OS Xenopus laevis (African clawed frog).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 CC Xenopodinae; Xenopus.  
 CX NCBI\_TaxID=8355;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Ovary;  
 RX MEDLINE=93181171; PubMed=8441611;  
 RA Rankin M.L., Heine M.A., Xiao S., Leblanc M.D., Nelson J.W.,  
 RA Dimario P.J.,  
 RT "A complete nucleolin cDNA sequence from Xenopus laevis";  
 RL Nucleic Acids Res. 21:169-169(1993).  
 [2]  
 RP SEQUENCE OF 125-650 FROM N.A.  
 RX MEDLINE=89252811; PubMed=2656405;  
 RA Calzergues-Ferrer M., Mariottini P., Curie C., Lapeyre B., Gas N.,  
 RA Analric F., Amaldi F.,  
 RT "Nucleolin from Xenopus laevis: cDNA cloning and expression during  
 development";  
 RL Genes Dev. 3:324-333(1989).  
 [3]  
 CC -!- FUNCTION: Nucleolin is the major nucleolar protein of growing  
 eukaryotic cells. It is found associated with intranucleolar  
 chromatin and peribosomal particles. It induces chromatin  
 decondensation by binding to histone H1. It is thought to play a  
 role in pre-rRNA transcription and ribosome assembly.  
 CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.  
 CC -!- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL; X63091; CAA44805.1; -;  
 PIR; S30250; S18874.  
 DR HSP; P11940; 1CVJ.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 4.  
 DR SMART; SM00360; RRM; 4.  
 DR PROSITE; PS0102; RRM; 4.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 3.  
 KW Nuclear protein; phosphorylation; Methylation; DNA-binding; Repeat;  
 RN RNA-binding.  
 FT INIT MET 0 BY SIMILARITY  
 FT DOMAIN 119 134 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 155 165 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 183 202 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 232 308 RNA-BINDING (RRM) 1.  
 FT DOMAIN 324 398 RNA-BINDING (RRM) 2.  
 FT DOMAIN 414 487 RNA-BINDING (RRM) 3.  
 FT DOMAIN 502 577 RNA-BINDING (RRM) 4.  
 FT DOMAIN 583 643 ARG/GLY/PHE-RICH.  
 FT MOD RES 154 154 PHOSPHORYLATION (BY SIMILARITY).  
 FT CONFLICT 214 214 P -> Q (IN REF. 2).  
 FT CONFLICT 218 219 PE -> LR (IN REF. 2).  
 FT CONFLICT 410 410 E -> Q (IN REF. 2).

FT CONFLICT 580 580 D -> E (IN REF. 2).  
 SQ SEQUENCE 650 AA; 70064 MW; 5BP9BF03768E71B4 CRC64;  
 Query Match 43.0%; Score 83; DB 1; Length 650;  
 Best Local Similarity 54.8%; Pred. No. 0.1;  
 Matches 17; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 2 KGGAATALEINLEEDDDNEDEVDQAEQ 32

Db 174 KGKTAQAEEEDDDDEDDDEEDDEEQ 204

## RESULT 3

SDC CAEEL STANDARD; PRT; 288 AA.  
 AC P50605;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Probable syndecan precursor.  
 GN F57C7.3.  
 OS Caenorhabditis elegans.  
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 CC Rhabditidae; Peloderinae; Caenorhabditis.  
 CX NCBI\_TaxID=6239;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA White S.;  
 RL Submitted (F8B-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Cell surface proteoglycan that bears heparan sulfate  
 (BY similarity).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
 CC -!- SIMILARITY: Belongs to the syndecan proteoglycan family.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL; Z69646; CAA93474.1; -;  
 PIR; T22846; T22846.  
 DR WormPep; P57C7.3; CE05996.  
 DR InterPro; IPR003585; Neurxin-like.  
 DR InterPro; IPR001050; Syndecan.  
 DR Pfam; PF01034; Syndecan; 1.  
 DR SMART; SM00294; 4.lm; 1.  
 DR PROSITE; PS00964; SYNDECAN; 1.  
 KW Proteoglycan; Heparan sulfate; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 26 POTENTIAL.  
 FT CHAIN 27 288 PROBABLE SYNDECAN.  
 FT DOMAIN 27 231 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 232 252 POTENTIAL.  
 FT DOMAIN 253 288 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 69 69 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 71 71 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).  
 FT CARBOHYD 86 86 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).  
 FT CARBOHYD 214 214 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).  
 FT DOMAIN 142 161 ASP/GLU-RICH (HIGHLY ACIDIC).  
 FT DOMAIN 175 181 POLY THR.  
 FT SEQUENCE 288 AA; 30979 MW; 10F14F118541341F CRC64;

Query Match 35.2%; Score 69; DB 1; Length 288;  
 Best Local Similarity 40.6%; Pred. No. 1.8;  
 Matches 13; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

QY 2 KGGAATALEINLEEDDDNEDEVDQAEQ 33

Db 129 KKPAADEKINVEEDDEDEDEDEDEDEE 160

CC entities requires a license agreement (see [http://www.isb-sib.ch/announcement/](http://www.isb-sib.ch/announcement) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
CC      EMBL; X53390; CAA37469.1; -
DR      EMBL; X53461; CAA37548.1; -
DR      EMBL; X5687; CAI0016.1; -
DR      EMBL; BC042297; AAH42297.1; -.
DR      PIR; S09318; S09318.
DR      PIR; S18193; S18193.
DR      PDZ; IYK99; O4-DEC-02.
DR      TRANSFAC; T02900; -.
DR      Genew; HGNC:12511; UBPf.
MIM; 600673; -.
DR      GO; GO:0005730; C:nucleolus; TAS.
DR      GO; GO:0003701; F:rRNA polymerase I transcription factor activity; TAS.
DR      GO; GO:0006356; P:regulation of transcription from Pol I prom.; TAS.
DR      InterPro; IPRO00910; HMG_12_box.
DR      Pfam; PF00505; HMG_box; 5.
DR      SMART; SM00398; HMG; 6.
DR      PROSITE; PS01118; HMG_BOX_2; 6.
KW      Transcription regulation; DNA-binding; Activator; Nuclear protein;
Repeat; alternative splicing; 3D-structure.
FT      DNA BIND   112     180
                   HMG BOX 1.
FF      DNA BIND   196     264
                   HMG BOX 2.
FF      DNA BIND   298     362
                   HMG BOX 3.
FF      DNA BIND   407     475
                   HMG BOX 4.
FF      DNA BIND   482     549
                   HMG BOX 5.
FF      DNA BIND   568     634
                   HMG BOX 6.
FF      DOWAIN    675     764
                   ASP/GLU/SER-RICH (ACIDIC).
FF      VASPLIC   221     257
                   Missing (in isoform UBF2).
                               /FTId=VSP_002193.
SQ      SEQUENCE   764 AA;  89406 MW;  D4F0FBEB1B0E757D CRC64;

Query Match          34.2%; Score 66; DB 1; Length 764;
Best Local Similarity 40.0%; Pred. No. 7.6;
Matches 12; Conservative              7; Mismatches 11; Indels 0; Gaps 0;

Qy       4 KGATALEINLRBEDDDNEDVEDEQAQQOK 33
         | :|::|||:||| ||| ::|
Db       665 KSRRTTLQSKESEDEDDEDDEDDEEE 694

RESULT 5
UBF1_MOUSE STANDARD;             PRT;    765 AA.
AC      P25976;
DT      01-MAY-1992 (Rel. 22, Created)
DT      01-MAY-1992 (Rel. 22, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Nucleolar transcription factor 1 (Upstream binding factor 1) (UBF-1).
DS      UBF1 OR TCUFJF OR UBF1 OR UBF-1.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI TaxId=10090;
RN      [1]
RP      SEQUENCE FROM N.A. (UBF1 AND UBF2).
RX      MEDLINE=91367658; PubMed=1891354;
RA      Hisataka K., Nishimura T., Maeda Y., Hanada K.I., Song C.Z.,
RA      Muramatsu W.;
RT      "Cloning and structural analysis of cDNA and the gene for mouse
RT      transcription factor UBF."
RL      Nucleic Acids Res. 19:4631-4637(1991).
CC      !- FUNCTION: UBF recognizes the ribosomal RNA gene promoter and
CC      activates transcription mediated by RNA polymerase I through
CC      cooperative interactions with the species-specific factor SL1. It
CC      binds specifically to the upstream control element.
CC      !- SUBUNIT: Homodimer.
CC      !- SUBCELLULAR LOCATION: Nuclear.
CC      !- ALTERNATIVE PRODUCTS:
CC      Event-Alternative splicing; Named isoforms=2;
CC      Name=UBF1; Synonyms=Long;
```

RESULT 4

UBF1\_HUMAN STANDARD; PRT; 764 AA.

P17480;

01-AUG-1990 (Rel. 15, Created)

01-AUG-1990 (Rel. 15, Last sequence update)

15-MAR-2004 (Rel. 43, Last annotation update)

Nucleolar transcription factor 1 (Upstream binding factor 1) (UBF-1) (Autoantigen NOR-90).

UBF1 OR UBF1 OR UBF.

Homio sapiens (Human).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

NCBI\_TaxId=9606;

[1]

SEQUENCE FROM N.A. (ISOFORM UBF1).

MEDLINE=90231434; PubMed=2330041;

Jantzen H.M., Admon A., Bell S.P., Tjian R.;

"Nucleolar transcription factor HUBF contains a DNA-binding motif with homology to HMG proteins.";

Nature 344:830-836(1990).

[2]

SEQUENCE FROM N.A. (ISOFORM UBF2).

MEDLINE=92044316; PubMed=1940801;

Chan E.K.U., Imai H., Hamei J.C., Tan E.M.;

"Human autoantigen to RNA polymerase I transcription factor HUBF. Molecular identity of nucleolus organizer region autoantigen NOR-90 and ribosomal RNA transcription upstream binding factor.";

J. Exp. Med. 174:1239-1244(1991).

[3]

SEQUENCE FROM N.A. (ISOFORM UBF2).

TISSUE=Testis;

MEDLINE=22388257; PubMed=12477932;

Strausberg R.L., Feigold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stopleton M., Soares M.B., Bonaldo M.F., Casavani T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano J.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Parey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blackley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

!- FUNCTION: Recognizes the ribosomal RNA gene promoter and activates transcription mediated by RNA polymerase I through cooperative interactions with the species-specific factor SL1. It binds specifically to the upstream control element.

!- SUBUNIT: Homodimer.

!- SUBCELLULAR LOCATION: Nuclear.

!- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=2;

Name=UBF1; Synonyms=Long;

Isoid=p17480-1; Sequence=Displayed;

Name=UBF2; Synonyms=Short;

Isoid=p17480-2; Sequence=VSP 002193;

!- SIMILARITY: Contains 6 HMG box\_domains.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial purposes is allowed.

```
CC isoid=p25976-1; Sequence=Displayed;
CC Name=UBF2; Synonyms=Short;
CC isoid=p25976-2; Sequence=VSP_002194;
CC -!- SIMILARITY: Contains 6 HMG box domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X60831; CAA43222.1; -
CC PIR; S22314; S22314.
CC HSSP; P07155; 1HMF.
CC MGD; MGI:98512; Ubf.
CC GO; GO:0005730; C:nucleolus; IDA.
CC GO; GO:0005515; F:protein binding; IPI.
CC InterPro; IPR00910; HMG_12_box.
CC Pfam; PF00505; HMG_box; 5.
CC SMART; SM00398; HMG; 6.
CC PROSITE; PS01118; HMG_BOX_2; 6.
CC Transcription regulation; DNA-binding; Activator; Nuclear protein;
CC Repeat; Alternative splicing.
CC FT DNA_BIND 112 180 HMG_BOX 1.
CC FT DNA_BIND 196 264 HMG_BOX 2.
CC FT DNA_BIND 298 362 HMG_BOX 3.
CC FT DNA_BIND 407 475 HMG_BOX 4.
CC FT DNA_BIND 482 549 HMG_BOX 5.
CC FT DNA_BIND 568 634 HMG_BOX 6.
CC FT DOMAIN 675 765 ASP/GLU/SER-RICH (ACTIDIC).
CC FT VARSPFLIC 221 257 Missing (in isoform UBF2).
CC FT /FTID=VSP_002194.
CC SQ SEQUENCE 765 AA; 89509 MW; 006A20A7F968DB6A CRC64;

Query Match 34.2%; Score 66; DB 1; Length 765;
Best Local Similarity 36.7%; Pred.No. 7.6;
Matches 11; Conservative 10; Mismatches 9; Indels 0; Gaps 0;

QY 4 KGAAATALEINLEEDDDNEDEVEQAEQOK 33
Db 665 KSRRTTQKSESEHDDDEEDDEEEEEE 694

RESULT 6
EF1B CANAL STANDARD; PRT; 213 AA.
AC P78590;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Elongation factor 1-beta (EF-1-beta).
GN EF1B.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WO-1;
RX MEDLINE=97120881; PubMed=8961551;
RA Maneu Flores V., Cervera A.M., Martinez J.P., Gozalbo D.;
RT "Molecular cloning and characterization of a Candida albicans gene
RT (EFB1) coding for the elongation factor EF-1 beta.";
RL FEMS Microbiol. Lett. 145:157-162(1996).
CC -!- FUNCTION: EF-1-beta and EF-1-delta stimulate the exchange of GDP
CC bound to EF-1-alpha to GTP.
CC -!- SUBUNIT: EF-1 is composed of four subunits: alpha, beta, delta,
CC and gamma.
CC -!- SIMILARITY: Belongs to the EF-1-beta/EF-1-delta family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X96517; CAA65366.1; -
CC InterPro; IPR001326; EPI_BD.
CC Pfam; PF00736; EPIBD; 1.
CC PROSITE; PS00824; EPIBD_1; 1.
CC PROSITE; PS00825; EPIBD_2; 1.
CC KW Elongation factor; Protein biosynthesis.
CC SQ SEQUENCE 213 AA; 23479 MW; 5C1976ADD4DB22CC CRC64;

Query Match 33.9%; Score 65.5; DB 1; Length 213;
Best Local Similarity 45.9%; Pred.No. 2.4;
Matches 17; Conservative 5; Mismatches 4; Indels 11; Gaps 2;

QY 3 GKGAATALEINLEEDDD-----NEDEVEQAEQOK 33
Db 75 GSAAAAA-----EEEDDEDVDLFGSDDEVEDEAEKLG 106

RESULT 7
NUCL_RAT STANDARD; PRT; 712 AA.
ID NUCL_RAT
AC P13383;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nucleolin (Protein C23).
GN NCL OR NUC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90269607; PubMed=2347493;
RA Bourbon H.-M., Amalric F.;
RT "Nucleolin gene organization in rodents: highly conserved sequences
RT within three of the 13 introns.";
RL Gene 88:187-196(1990).
RN [2]
RP SEQUENCE OF 1-44 FROM N.A.
RX MEDLINE=89121496; PubMed=2906027;
RA Bourbon H.-M., Prudhomme M., Amalric F.;
RT "Sequence and structure of the nucleolin promoter in rodents:
RT characterization of a strikingly conserved CpG island.";
RL Gene 68:73-84(1988).
CC -!- FUNCTION: Nucleolin is the major nucleolar protein of growing
CC eukaryotic cells. It is found associated with intranuclear
CC chromatin and preribosomal particles. It induces chromatin
CC decondensation by binding to histone H1. It is thought to play a
CC role in pre-rRNA transcription and ribosome assembly.
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -!- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M55022; AAA41732.1; -
CC EMBL; M55015; AAA41732.1; JOINED.
CC EMBL; M55017; AAA41732.1; JOINED.
CC EMBL; M55020; AAA41732.1; JOINED.
CC EMBL; M22090; AAA41733.1; -
CC HSSP; P09651; 1HAI.
```

```

R InterPro: IPR000504; RNA_rec_mot.
R Pfam: PF00076; rrm; 4.
R SMART: SW00360; RRM; 4.
R R PROSITE: PS0102; RRM; 4.
R R PROSITE: PS0030; RRM RNP 1; 3.
R W Nuclear protein; Phosphorylation; Methylation; DNA-binding; Repeat;
W RNA-binding.
T INIT MET 0
T DOMAIN 142 167 ASP/GLU-RICH (ACIDIC)
T DOMAIN 187 215 ASP/GLU-RICH (ACIDIC)
T DOMAIN 241 274 ASP/GLU-RICH (ACIDIC)
T DOMAIN 310 386 RNA-BINDING (RRM) 1.
T DOMAIN 396 469 RNA-BINDING (RRM) 2.
T DOMAIN 488 562 RNA-BINDING (RRM) 3.
T DOMAIN 574 649 RNA-BINDING (RRM) 4.
T DOMAIN 651 702 ARG/GLY/PHE-RICH.
T DOMAIN 57 134
T REPEAT 57 64
T REPEAT 74 81
T REPEAT 82 89
T REPEAT 90 97
T REPEAT 98 103
T REPEAT 104 111
T REPEAT 119 126
T REPEAT 127 134
T REPEAT 127 134
T SEQUENCE 712 AA; 77016 MW; 68774A214E550F90 CRC64;

Query Match 33.7%; Score 65; DB 1; Length 712;
Best Local Similarity 40.6%; Pred. No. 9;
Matches 13; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

Y 2 KGKGAATALEINLEEDDNEDEVDQAEQKTHV 33
b 178 KAAPAPASEDEDEDDDDDDDEDEEE 209

RESULT 8
S32_DICDI
D HS32_DICDI STANDARD; PRT; 281 AA.
C P5458;
T 01-OCT-1996 (Rel. 34, Created)
T 01-OCT-1996 (Rel. 34, Last sequence update)
T 15-MAR-2004 (Rel. 43, Last annotation update)
E 32 kDa heat shock protein (4-1 protein).
N HSPC OR HSP32 OR DD 00380.
S Dictyostelium discoideum (Slime mold).
C Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
N NCBI_TaxID=44689;
X MEDLINE=96001265; PubMed=7557471;
X de Maria A.C., Gomes S.L., Juliania M.H., Mazzarella R., Klein C.;
T "Cloning of a cDNA encoding a novel heat-shock protein from
T Dictyostelium discoideum."
L Gene 163:163-164(1995).
N [1]
P SEQUENCE FROM N.A.
X MEDLINE=22092622; PubMed=12097910;
X Gloeckner G., Eichinger L., Szafranski K., Pachbat J.A.,
A Bankier A.T., Dear P.H., Lehmann R., Baumgart C., Parra G.,
A Abril J.F., Guigo R., Kumpf K., Tunggal B., Cox E., Quail M.A.,
A Platzer M., Rosenthal A., Noegel A.A.;
T "Sequence and analysis of chromosome 2 of Dictyostelium discoideum."
L Nature 418:79-85(2002).
C -!- DEVELOPMENTAL STAGE: Present at high levels in growing cells but
C decreases dramatically during the early hours of development.
C -!- INDUCTION: By heat shock.
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its

```

---

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; U40211; AAC47710.1; -.
DR ENBL; L39778; AAA99510.1; -.
DR ENBL; AC116955; AAM44326.1; -.
DR ENBL; AC117176; AAO52101.1; -.
DR PIR; JC4295; JC4295.
DR DictyBase; DDB0185048; hspC.
KW Heat shock.
FT DOMAIN 251 257 POLY-GLN.
FT DOMAIN 262 265 POLY-ASN.
SQ SEQUENCE 281 AA; 31481 MW; 2C64F57C847D7CA4 CRC64;

Query Match 33.4%; Score 64.5; DB 1; Length 281;
Best Local Similarity 42.9%; Pred. No. 4.1;
Matches 15; Conservative 8; Mismatches 11; Indels 1; Gaps 1;

QY 2 KGKGAATALEINLEEDDNEDEVDQAEQKTHV 36
Db 39 KGKYLTV-ISLWEDEMEEDDDVDEERSPREDI 72

RESULT 9
YI97_HUMAN
ID YI97_HUMAN STANDARD; PRT; 661 AA.
AC Q96PZ0; Q9NX19;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical pseudouridine synthase KIAA1997 (EC 4.2.1.70).
GN KIAA1997.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye, and Urinary bladder;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Heiton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE OF 23-661 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21456161; PubMed=11572484;
RA Nagase T., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XXI.
RT The complete sequences of 60 new cDNA clones from brain which code for
RT large proteins."
RL DNA Res. 8:179-187(2001).
RN [3]
RP SEQUENCE OF 420-661 FROM N.A.

```

```
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isoigai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
CC Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine
CC 5'-phosphate + H(2)O.
CC -!- SIMILARITY: Belongs to the pseudouridine synthase truD family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: BC005209; AA05209.2; ALT_INIT.
CC EMBL: BC011396; AA011396.1; ALT_INIT.
CC EMBL: AB067484; BA067790.1; -.
CC EMBL: AK000492; BA091203.1; ALT_INIT.
CC InterPro: IPR001656; UPF0024.
CC Pfam: PF01142; UPF0024; 1.
CC TIGRFAMs: TIGR00094; TIGR00094; 1.
CC KW Hypothetical protein; tRNA processing; Lyase.
CC FT ACT SITE 294 294 BY SIMILARITY.
CC FT CONFLICT 244 244 N -> KVTATAD (IN REF. 2).
CC SQ SEQUENCE 661 AA; 75035 MW; 6F6A05A9B57B1560 CRC64;
CC -----
Query Match 33.4%; Score 64.5; DB 1; Length 661;
Best Local Similarity 34.3%; Pred. No. 9.5;
Matches 12; Conservative 11; Mismatches 11; Indels 1; Gaps 1;
CC
QY 3 GKGAATALEINLEEDDDNDEVDQAEQKTHVF 37
DB 68 GKGGKNS-EAQLEDEEEEDPGLSECEESRSF 101
CC -----
RESULT 10
ID DEMA_MOUSE STANDARD; PRT; 405 AA.
AC Q9WV69;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Dectin (Erythrocyte membrane protein band 4.9).
GN EPB49 OR EPB4.9.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=99431674; PubMed=10501976;
RA Azim A.C., Kim A.C., Lutchman M., Andrabi S., Peters L.L.,
RA Chishti A.H.;
RT "cDNA sequence, genomic structure, and expression of the mouse dematin
RT gene [EPB4.9].";
RL Mamm. Genome 10:1026-1029(1999).
CC -!- FUNCTION: Actin-bundling protein (By similarity).
CC -!- DOMAIN: Contains at least two actin-binding sites, one in the
CC headpiece domain and one in the amino-terminal portion.
CC -!- DOMAIN: Consists of a large core fragment, the amino-terminal
CC portion, and a small headpiece, the carboxyl-terminal portion. The
CC headpiece can bind but cannot bundle actin filaments.
CC -!- SIMILARITY: Belongs to the villin/gelsolin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF155547; AAD38412.1; -.
CC HSSP; P02640; 1VII.
CC MGD; MGI:99670; EPB4.9.
CC InterPro: IPR003128; VHP.
CC Pfam: PF02209; VHP; 1.
CC SMART; SM00153; VHP; 1.
CC KW Actin-binding; Repeat; Phosphorylation; Actin capping.
CC FT DOMAIN 216 222 POLY-GLU.
CC FT DOMAIN 318 405 HEADPIECE (BY SIMILARITY).
CC FT MOD RES 403 403 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
CC SQ SEQUENCE 405 AA; 45468 MW; AECAS52500BDD19A CRC64;
CC -----
Query Match 33.2%; Score 64; DB 1; Length 405;
Best Local Similarity 43.8%; Pred. No. 6.6;
Matches 14; Conservative 5; Mismatches 13; Indels 0; Gaps 0;
CC
QY 2 KGGAATALEINLEEDDDNDEVDQAEQK 33
DB 207 KKKASRKGAEEEEEDDDSEEEKAIRQK 238
CC -----
RESULT 11
ID AMP2_RAT STANDARD; PRT; 478 AA.
AC P38062;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Methionine aminopeptidase 2 (EC 3.4.11.18) (Metap 2) (Peptidase M 2)
DE (Initiation factor 2 associated 67 kDa glycoprotein) (p67) (p67eIF2).
DE METAP2 OR MNPEP OR P67EIF2.
DE Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Reuber H35; TISSUE=Liver;
RX MEDLINE=93266517; PubMed=8496145;
RA Wu S., Gupta S., Chatterjee N., Hileman R.E., Kinzy T.G., Gupta N.K.;
RA Denslow N.D., Merrick W.C., Chakrabarti D., Osterman J.C., Gupta N.K.;
RT "Cloning and characterization of complementary DNA encoding the
RT eukaryotic initiation factor 2-associated 67-kDa protein (p67).";
RL J. Biol. Chem. 268:10796-10801(1993).
RN [2]
RP REVISIONS TO C-TERMINUS.
RX MEDLINE=95372350; PubMed=7644482;
RA Arfin S.M., Kendall R.L., Hall L., Weaver L.H., Stewart A.E.,
RA Matthews B.W., Bradshaw R.A.;
RT "Eukaryotic methionyl aminopeptidases: two classes of
RT cobalt-dependent enzymes.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:7714-7718(1995).
CC -!- FUNCTION: Removes the amino-terminal methionine from nascent
CC proteins.
CC -!- FUNCTION: PROTECTS EIF-2 ALPHA-SUBUNIT FROM INHIBITORY
CC PHOSPHORYLATION BY EIF-2 KINASES. PLAYS A CRITICAL ROLE IN THE
CC REGULATION OF PROTEIN SYNTHESIS. IT ALSO INTERACTS WITH THE
CC EIF-2 GAMMA-SUBUNIT.
CC -!- CATALYTIC ACTIVITY: Release of N-terminal amino acids.
CC Preferentially methionine, from peptides and arylamides.
CC -!- COFACTOR: Cobalt; binds 2 ions per subunit (By similarity).
CC -!- PTM: CONTAINS 12 O-LINKED GLNAC.
CC -!- SIMILARITY: Belongs to peptidase family M24C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```



or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)

EMBL; L10652; AAA41111.1; -.  
PIR; A46702; A46702.

HSSP; P50579; 1B6A.

MEROPS; M24.002; --

InterPro; IPR001714; Pept\_M2

InterPro; IPR002468; Pept\_M2

InterPro; IPR000994; Peptida

Prism; PFO55/; Peptidease\_M24

PRINIS; PRO0399; MAPEPIIDASE  
TIGREAMS; TIGR00501: met nda

PROSTATE: PS01202: MAP 2: 1

Hydrolase: Aminonitrilase: C

DOMAIN	36	46
...	...	...

DOMAIN	98	106
--------	----	-----

METAL 251 251

**METAL** 262 262

METAL 331 331

**METAL** 364 364

METAL	459	459
-------	-----	-----

CONFLICT 464 478

3  
 6  
 1  
 5  
 1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522

SEQUENCE 478 AA; 53052 MW

32 38.

est Local similarity 45.2%;

14: Conservative

0  
.  
4  
0  
3  
.  
4  
0  
2  
4  
0  
0  
.  
4  
4  
  
0  
0  
4  
0  
0  
1

2 KGKGAAATALEINLEEEEDDD

[illegible]

44 KGKGAVSAGQQQLDKESGT

JUL 12 1964

RA Akusjaervi G., Alestroem P., Pettersson M., Lager M., Joernvall H.,  
RA Pettersson U.;  
RT "The gene for the adenovirus 2 hexon polypeptide.";  
RL J. Biol. Chem. 259:13976-13979(1984).  
RN [2]  
RP SEQUENCE OF 1-454 AND 456-967.  
RX MEDLINE=81215564; PubMed=6263309;  
RA Joernvall H., Akusjaervi G., Alestroem P., von Bahr-Lindstroem H.,  
RA Pettersson U., Appella E., Fowler A.V., Philipson L.;  
RT "The adenovirus hexon protein. The primary structure of the  
RT polypeptide and its correlation with the hexon gene.";  
RL J. Biol. Chem. 256:6181-6186(1981).  
RN [3]  
RP SEQUENCE OF 922-966 FROM N.A.  
RX MEDLINE=81150446; PubMed=6259616;  
RA Akusjaervi G., Zabieliski J., Perricaudet M., Pettersson U.;  
RT "The sequence of the 3' non-coding region of the hexon mRNA discloses  
RT a novel adenovirus gene.";  
RL Nucleic Acids Res. 9:1-17(1981).  
RN [4]  
RP ACETYLATION  
RX MEDLINE=74147480; PubMed=4823869;  
RA Joernvall H., Ohlsson H., Philipson L.;  
RT "An acetylated N-terminus of adenovirus type 2 hexon protein.";  
RL Biochem. Biophys. Res. Commun. 56:304-310(1974).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).  
RX MEDLINE=95019210; PubMed=7932702;  
RA Athappilly F.K., Murail R., Rux J.J., Cai Z., Burnett R.M.;  
RT "The refined crystal structure of hexon, the major coat protein of  
RT adenovirus type 2, at 2.9-A resolution.";  
RL J. Mol. Biol. 242:430-455(1994).  
CC -!- FUNCTION: This protein is one of the structural proteins in the  
CC viral coat and is synthesized during late infection.  
CC -!- SUBUNIT: Homotrimer.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; J01917; AAA92215.1; -;  
DR PIR; A94597; HXAD2.  
DR PDB; 1P2Z; 11-NOV-03.  
DR InterPro; IPR000736; Adeno\_hexon.  
DR Pfam; PF01065; Adeno\_hexon\_1.  
DR Pfam; PF03678; Adeno\_hexon\_C; 1.  
DR ProDom; PD002815; Adeno\_hexon; 1.  
KW Coat protein; Hexon protein; Late protein; Acetylation; 3D-structure.  
FT INIT MET 0 0  
FT MOD\_RES 1 1 ACETYLATION.  
FT FT DOMAIN 133 161 GLU-RICH (ACIDIC).  
FT FT HELIX 50 53  
FT FT TURN 54 54  
FT FT TURN 60 63  
FT FT STRAND 98 98  
FT FT STRAND 101 101  
FT FT STRAND 112 112  
FT FT STRAND 119 123  
FT FT TURN 125 126  
FT FT STRAND 130 130  
FT FT HELIX 131 133  
FT FT STRAND 135 143  
FT FT TURN 146 147  
FT FT STRAND 166 167  
FT FT STRAND 170 170  
FT FT TURN 186 188  
FT FT STRAND 215 215  
FT FT STRAND 222 222  
FT FT TURN 228 230  
FT FT TURN 234 234  
FT FT STRAND 235 235  
FT FT STRAND 247 247  
FT FT TURN 253 253  
FT FT STRAND 267 267  
FT FT STRAND 289 289  
FT FT STRAND 305 305  
FT FT STRAND 309 313  
FT FT TURN 317 318  
FT FT HELIX 323 327  
FT FT STRAND 329 332  
FT FT STRAND 339 340  
FT FT STRAND 342 344  
FT FT STRAND 345 345  
FT FT TURN 346 347  
FT FT STRAND 348 348  
FT FT STRAND 355 358  
FT FT TURN 362 363  
FT FT TURN 369 370  
FT FT TURN 372 374  
FT FT HELIX 377 388  
FT FT STRAND 395 395  
FT FT HELIX 396 398  
FT FT TURN 399 399  
FT FT STRAND 400 400  
FT FT STRAND 407 410  
FT FT STRAND 413 413  
FT FT STRAND 423 425  
FT FT TURN 428 429  
FT FT STRAND 435 437  
FT FT TURN 448 449  
FT FT TURN 453 454  
FT FT STRAND 465 467  
FT FT STRAND 475 477  
FT FT HELIX 479 489  
FT FT TURN 490 492  
FT FT HELIX 493 495  
FT FT STRAND 498 500  
FT FT STRAND 502 502  
FT FT TURN 515 515  
FT FT HELIX 516 521  
FT FT STRAND 523 523  
FT FT TURN 526 527  
FT FT TURN 531 536  
FT FT STRAND 538 538  
FT FT TURN 542 543  
FT FT HELIX 544 546  
FT FT TURN 549 550  
FT FT TURN 553 554  
FT FT HELIX 556 563  
FT FT TURN 564 565  
FT FT STRAND 570 573  
FT FT STRAND 576 577  
FT FT TURN 582 585  
FT FT STRAND 586 586  
FT FT STRAND 588 588  
FT FT STRAND 592 597  
FT FT STRAND 600 601  
FT FT HELIX 604 607  
FT FT TURN 616 620  
FT FT STRAND 624 631  
FT FT STRAND 633 633  
FT FT TURN 640 641  
FT FT HELIX 642 645  
FT FT STRAND 646 648  
FT FT STRAND 654 657  
FT FT STRAND 663 669  
FT FT STRAND 675 677  
FT FT STRAND 682 683  
FT FT TURN 685 686  
FT FT STRAND 689 690  
FT FT STRAND 695 697



Search completed: March 7, 2004, 13:35:44  
Job time : 10.3516 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

WM protein - protein search, using sw model

Run on: March 7, 2004, 13:31:21 ; Search time 40.2527 Seconds

(without alignments)  
290.022 Million cell updates/sec

Title: US-09-643-458B-9

Perfect score: 193

Sequence: 1 CKGKAATALEINLEEDDDNEVDEQAQQKTHVF 37

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL.25.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriaph.\*
- 17: sp\_archesp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	167	86.5	952	12 Q805J1	Q805j1 human adeno
2	164	85.0	952	12 Q80RJ2	Q80rj2 human adeno
3	164	85.0	952	12 Q80R10	Q80r10 human adeno
4	159	82.4	952	12 Q80R15	Q80r15 human adeno
5	72.5	37.6	983	5 Q81DY7	Q81dy7 plasmodium
6	70	36.3	612	10 Q9LGT1	Q9lgt1 oryza sativ
7	68	35.2	287	5 Q814H9	Q814h9 caenorhabdi
8	68	35.2	1188	11 P97500	P97500 mus musculu
9	68	35.2	1211	11 Q8CHB4	Q8chb4 mus musculu
10	67	34.7	270	11 Q80VC5	Q80vc5 mus musculu
11	67	34.7	380	11 Q8BZV6	Q8bzv6 mus musculu
12	67	34.7	380	11 Q7TNV0	Q7tnv0 mus musculu
13	66.5	34.5	396	10 Q9SDZ5	Q9sdz5 oryza sativ
14	66	34.2	619	11 Q8VCZ7	Q8vcz7 mus musculu
15	66	34.2	654	4 O00164	O00164 homo sapien
16	65.5	33.9	371	10 Q9SK74	Q9sk74 arabidopsis

#### ALIGNMENTS

RESULT 1

ID Q805J1 PRELIMINARY; PRT; 952 AA.  
AC Q805J1;  
DT 01-JUN-2003 (TREMBLrel. 24, Created)  
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Hexon.  
OS Human adenovirus type 5.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
OX NCBI\_taxid=28285;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KNIH Ad00/12, KNIH Ad00/19, and KNIH Ad01/1;  
RA Na B.-K., Lee J.-Y., Kim J.-H., Kim W.-J., Kang C.;  
RT "Sequence analysis of hexon protein of adenoviruses isolated in Korea";  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF542124; AAC24099.1; -;  
DR EMBL; AF542128; AAC24103.1; -;  
DR EMBL; AF542130; AAC24105.1; -;  
DR GO; GO:0019028; C: viral capsid; IEA.  
DR GO; GO:0005198; P: structural molecule activity; IEA.  
DR InterPro; IPR000736; Adeno\_hexon.  
DR Pfam; PF01065; Adeno\_hexon; 1.  
DR Pfam; PF03678; Adeno\_hexon; 1.  
DR ProDom; PD002815; Adeno\_hexon; 1.  
SQ SEQUENCE 952 AA; 107914 MW; C92738D3C0383802 CRC64;

Query Match 86.5%; Score 167; DB 12; Length 952;  
Best Local Similarity 89.2%; Pred. No. 6.1e-12;  
Matches 33; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CKGKAATALEINLEEDDDNEVDEQAQQKTHVF 37

Db 133 CEWDEAATALEINLEEDDDNEVDEQAQQKTHVF 169

RESULT 2

Q91029 human adeno  
Q99K50 mus musculu  
P70475 rattus norv  
Q7ZV24 xenopus lae  
Q9M0S8 arabidopsis  
Q8AVU3 xenopus lae  
Q96UB5 neurospora  
Q8JZV5 mus musculu  
Q9PSW7 xenopus lae  
Q9WM2 mus musculu  
Q9CT46 mus musculu  
Q9UBP6 clona intes  
Q8BD8 mus musculu  
Q8140 plasmodium  
Q8CE30 mus musculu  
Q8CD23 mus musculu  
Q80BM9 salmeline  
O45181 caenorhabdi  
Q912J7 human adeno  
Q91028 human adeno  
Q80I13 human adeno  
Q80I11 human adeno  
Q7TEH9 human adeno  
Q84N08 oryza sativ  
Q9P4C6 pichia past  
O08996 mus musculu  
Q81I24 plasmodium  
Q80YZ1 mus musculu

Q80RJ2 PRELIMINARY; PRT; 952 AA.  
ID Q80RJ2 AC Q80RJ2  
DT 01-JUN-2003 (TRENBLrel. 24, Created)  
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Hexon.  
OS Human adenovirus type 5.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
OX NCBI\_TaxID=28285;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KNH Ad99/5;  
RA Na B.-K., Lee J.-Y., Kim J.-H., Kim W.-J., Kang C.;  
RT "Sequence analysis of hexon protein of adenoviruses isolated in Korea."  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF542116; AAC24084.1; -;  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR InterPro; IPR000736; F:structural molecule activity; IEA.  
DR Pfam; PF01065; Adeno\_hexon.1.  
DR Pfam; PF03678; Adeno\_hexon.C; 1.  
DR ProDom; PD002815; Adeno\_hexon; 1.  
SQ SEQUENCE 952 AA; 107938 MW; E95940D564209EAB CRC64;  
Query Match 85.0%; Score 164; DB 12; Length 952;  
Best Local Similarity 86.5%; Pred. No. 1.4e-11;  
Matches 32; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 1 CKGKAATALEINLEEDDDNEDEVDQAEQKTHVF 37  
DB 133 CEWDEAATALENNLEEDDDNEDEVDQAEQKTHVF 169  
RESULT 3  
Q80RI0 PRELIMINARY; PRT; 952 AA.  
ID Q80RI0 AC Q80RI0  
DT 01-JUN-2003 (TRENBLrel. 24, Created)  
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Hexon.  
OS Human adenovirus type 5.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
OX NCBI\_TaxID=28285;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KNH Ad00/5;  
RA Na B.-K., Lee J.-Y., Kim J.-H., Kim W.-J., Kang C.;  
RT "Sequence analysis of hexon protein of adenoviruses isolated in Korea."  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF542121; AAC24096.1; -;  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR InterPro; IPR000736; F:structural molecule activity; IEA.  
DR Pfam; PF01065; Adeno\_hexon.1.  
DR Pfam; PF03678; Adeno\_hexon.C; 1.  
DR ProDom; PD002815; Adeno\_hexon; 1.  
SQ SEQUENCE 952 AA; 107972 MW; 35440D68DC58E71C CRC64;  
Query Match 85.0%; Score 164; DB 12; Length 952;  
Best Local Similarity 86.5%; Pred. No. 1.4e-11;  
Matches 32; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 1 CKGKAATALEINLEEDDDNEDEVDQAEQKTHVF 37  
DB 133 CEWDEAATALENNLEEDDDNEDEVDQAEQKTHVF 169  
RESULT 4  
Q80RI5 PRELIMINARY; PRT; 952 AA.  
ID Q80RI5 AC Q80RI5  
DT 01-JUN-2003 (TRENBLrel. 24, Created)  
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Hexon.  
OS Human adenovirus type 5.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
OX NCBI\_TaxID=28285;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KNH Ad99/12;  
RA Na B.-K., Lee J.-Y., Kim J.-H., Kim W.-J., Kang C.;  
RT "Sequence analysis of hexon protein of adenoviruses isolated in Korea."  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF542116; AAC24091.1; -;  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR InterPro; IPR000736; F:structural molecule activity; IEA.  
DR Pfam; PF01065; Adeno\_hexon.1.  
DR Pfam; PF03678; Adeno\_hexon.C; 1.  
DR ProDom; PD002815; Adeno\_hexon; 1.  
SQ SEQUENCE 952 AA; 107915 MW; 0514B635D6CADA54 CRC64;  
Query Match 82.4%; Score 159; DB 12; Length 952;  
Best Local Similarity 83.8%; Pred. No. 5.8e-11;  
Matches 31; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 1 CKGKAATALEINLEEDDDNEDEVDQAEQKTHVF 37  
DB 133 CEWDEAATALENNLEEDDDNEDEVDQAEQKTHVF 169  
RESULT 5  
Q8IDY7 PRELIMINARY; PRT; 983 AA.  
ID Q8IDY7 AC Q8IDY7  
DT 01-MAR-2003 (TRENBLrel. 23, Created)  
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN PF13\_0189.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,  
RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,  
RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrrell B.;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL844509; CAD52480.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 983 AA; 117789 MW; 28A71940943628F3 CRC64;  
Query Match 37.6%; Score 72.5; DB 5; Length 983;  
Best Local Similarity 45.0%; Pred. No. 2.1;  
Matches 18; Conservative 8; Mismatches 7; Indels 7; Gaps 2;  
QY 3 GKGATALEINLEEDDDNEDEVDQAEQKTHVF 35  
DB 745 GKLDATPIKHYKFEQDDDDNEEVEDEEDDEAENQKNY 784  
RESULT 6  
Q9LGT1 PRELIMINARY; PRT; 612 AA.  
ID Q9LGT1 AC Q9LGT1  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
DE P0489A01.10 protein.

Q80RI5 PRELIMINARY; PRT; 952 AA.  
ID Q80RI5 AC Q80RI5  
DT 01-JUN-2003 (TRENBLrel. 24, Created)  
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Hexon.  
OS Human adenovirus type 5.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
OX NCBI\_TaxID=28285;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KNH Ad99/12;  
RA Na B.-K., Lee J.-Y., Kim J.-H., Kim W.-J., Kang C.;  
RT "Sequence analysis of hexon protein of adenoviruses isolated in Korea."  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF542116; AAC24091.1; -;  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR InterPro; IPR000736; F:structural molecule activity; IEA.  
DR Pfam; PF01065; Adeno\_hexon.1.  
DR Pfam; PF03678; Adeno\_hexon.C; 1.  
DR ProDom; PD002815; Adeno\_hexon; 1.  
SQ SEQUENCE 952 AA; 107915 MW; 0514B635D6CADA54 CRC64;  
Query Match 82.4%; Score 159; DB 12; Length 952;  
Best Local Similarity 83.8%; Pred. No. 5.8e-11;  
Matches 31; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 1 CKGKAATALEINLEEDDDNEDEVDQAEQKTHVF 37  
DB 133 CEWDEAATALENNLEEDDDNEDEVDQAEQKTHVF 169  
RESULT 5  
Q8IDY7 PRELIMINARY; PRT; 983 AA.  
ID Q8IDY7 AC Q8IDY7  
DT 01-MAR-2003 (TRENBLrel. 23, Created)  
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN PF13\_0189.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,  
RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,  
RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrrell B.;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL844509; CAD52480.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 983 AA; 117789 MW; 28A71940943628F3 CRC64;  
Query Match 37.6%; Score 72.5; DB 5; Length 983;  
Best Local Similarity 45.0%; Pred. No. 2.1;  
Matches 18; Conservative 8; Mismatches 7; Indels 7; Gaps 2;  
QY 3 GKGATALEINLEEDDDNEDEVDQAEQKTHVF 35  
DB 745 GKLDATPIKHYKFEQDDDDNEEVEDEEDDEAENQKNY 784  
RESULT 6  
Q9LGT1 PRELIMINARY; PRT; 612 AA.  
ID Q9LGT1 AC Q9LGT1  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
DE P0489A01.10 protein.

```

N P0489A01.10.
S Oryza sativa (Rice).
C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
C Ehrhartoideae; Oryzeae; Oryza.
X NCBI_TaxID=4530;
N (1)
P SEQUENCE FROM N.A.
C STRAIN=cv. Nipponbare;
P "Oryza sativa matsumoto T., Yamamoto K.;
T "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
L clone:P0489A01."
L Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
R EMBL; AP002484; BAA9519.1; -.
R Gramene; Q9LGT1;
R GO; GO:0004437; F:inositol/phosphatidylinositol phosphatase a. . .; IEA.
R InterPro; IPR005135; Exo_endo_phos.
R InterPro; IPR000300; IPPC.
R Pfam; PF03372; Exo_endo_phos; 1.
R SMART; SM00128; IPPC; 1.
Q SEQUENCE 612 AA; 67565 MW; A65DCC31D9764A54 CRC64;

Query Match 36.3%; Score 70; DB 10; Length 612;
Best Local Similarity 65.0%; Pred. No. 2.6;
Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Y 11 EINLEEDDDNEDEVDQAE 30
b 32 EFNADEEDDDGSDVDQAE 51

RESULT 7
D Q814H9 PRELIMINARY; PRT; 287 AA.
C Q814H9;
T 01-MAR-2003 (TrEMBLrel. 23, Created)
T 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
E C. elegans sdh-1 protein (corresponding sequence F57C7.3b).
N F57C7.3 OR SDN-1.
S Caenorhabditis elegans.
C Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
C Rhabditidae; Peloderinae; Caenorhabditis.
X NCBI_TaxID=6239;
N (1)
P SEQUENCE FROM N.A.
C White S.;
L Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
N (2)
P SEQUENCE FROM N.A.
X MEDLINE=99069613; PubMed=9851916;
A none;
T "Genome sequence of the nematode C.elegans: A platform for
T investigating biology."
L Science 282:2012-2018(1998).
R EMBL; Z69646; CAD44142.1; -.
R WormPep; F57C7.3b; CE32581.
R GO; GO:0016020; C:membrane; IEA.
R GO; GO:0008092; F:cytoskeletal protein binding; IEA.
R InterPro; IPR001050; Syndecan.
R Pfam; PF01034; Syndecan; 1.
R PROSITE; PS00964; SYNDECAN; 1.
Q SEQUENCE 287 AA; 30851 MW; C947CF0A7FC95BC5 CRC64;

Query Match 35.2%; Score 68; DB 5; Length 287;
Best Local Similarity 40.6%; Pred. No. 2.2;
Matches 13; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

Y 2 KGKGAATALEINLEEDDDNEDEVDQAEQQK 33
b 128 KKPAPANDKEIKVEDEDDDEDEDEDEDEDE 159

```

```

RESULT 8
P97500
ID P97500 PRELIMINARY; PRT; 1188 AA.
AC P97500;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Zinc finger protein Png-1.
GN MYTIL OR PNG-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Brain;
RX MEDLINE=97276971; PubMed=9130664;
RA Weiner J.A., Chun J.;
RT "Png-1, a nervous system-specific zinc finger gene, identifies regions
RT containing postmitotic neurons during mammalian embryonic
RT development."
RL J. Comp. Neurol. 381:130-142(1997).
DR EMBL; U86338; AAC53157.1; -.
DR FIR; T46608; T46608.
DR MGD; MGI:1100511; Mytil.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR002515; Znf_C2HC.
DR Pfam; PF01530; zf-C2HC; 6.
SQ SEQUENCE 1188 AA; 133000 MW; D8D81C851C50E59E CRC64;

Query Match 35.2%; Score 68; DB 11; Length 1188;
Best Local Similarity 44.0%; Pred. No. 8.8;
Matches 11; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Qy 11 EINLEEDDDNEDEVDQAEQQKXTH 35
Db 151 DVVEEDDDDEEEEEEEDH 175

RESULT 9
Q8CHB4 PRELIMINARY; PRT; 1211 AA.
ID Q8CHB4
AC Q8CHB4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE KIAA1106 protein (Fragment).
GN KIAA1106
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Hara Y., Nagase T.,
RA Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT 1. The complete nucleotide sequences of 100 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries."
RT Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AB093283; BAC41467.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR002515; Znf_C2HC.
DR Pfam; PF01530; zf-C2HC; 6.
DR NON TER 1
FT SEQUENCE 1211 AA; 135470 MW; 36880EE2191AFCE4 CRC64;
SQ

```

Query Match 35.2%; Score 68; DB 11; Length 1211;  
 Best Local Similarity 44.0%; Pred. No. 9;  
 Matches 11; Conservative 7; Mismatches 0; Gaps 0;

QY 11 E1NLEEDDDNDEVDQAEQOKTH 35  
 : : ||||| : : : : :  
 Db 176 DVEEEDDDDEEEEEEENEDH 200

RESULT 10

Q80VC5 PRELIMINARY; PRT; 270 AA.

AC Q80VC5;  
 DT 01-JUN-2003 (TREMELrel. 24, Created)  
 DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE Similar to RIKEN cdna 1810019E15 gene (Fragment).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CZECH II; TISSUE=Breast tumor;  
 RA Strausberg R.;  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC048844; AAH48844.1; -;  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR InterPro; IPR003034; SAP.  
 DR Pfam; PF02037; SAP; 1.  
 DR SMART; SM00513; SAP; 1.  
 FT NON TER 270 270  
 FT SEQUENCE 270 AA; 30680 MW; 4EF095E81D4ABEBE CRC64;

Query Match 34.7%; Score 67; DB 11; Length 270;  
 Best Local Similarity 55.0%; Pred. No. 2.7;  
 Matches 11; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

QY 15 EEEDDDNDEVDQAEQOKT 34  
 ||||| : : : : :  
 Db 36 EEEDDDDEDEEEKEKS 55

RESULT 11

Q8BZV6 PRELIMINARY; PRT; 380 AA.

AC Q8BZV6;  
 DT 01-WAR-2003 (TREMELrel. 23, Created)  
 DT 01-WAR-2003 (TREMELrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE DEK protein homolog.  
 GN DEK OR 1810019E15RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Colon;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium.  
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs."  
 RL Nature 420:563-573 (2002).  
 DR EMBL; AK033451; BAC28296.1; -;  
 DR MGD; MGI:1926209; Dek.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR InterPro; IPR003034; SAP.  
 DR Pfam; PF02037; SAP; 1.  
 DR SMART; SM00513; SAP; 1.

SQ SEQUENCE 380 AA; 43098 MW; CE1F111076CE5AC1 CRC64;  
 Query Match 34.7%; Score 67; DB 11; Length 380;  
 Best Local Similarity 55.0%; Pred. No. 3.8;  
 Matches 11; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

QY 15 EEEDDDNDEVDQAEQOKT 34  
 ||||| : : : : :  
 Db 36 EEEDDDDEDEEEKEKS 55

RESULT 12

Q7TNV0 PRELIMINARY; PRT; 380 AA.

ID Q7TNV0;  
 DT 01-OCT-2003 (TREMELrel. 25, Created)  
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Salivary gland;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L.; Feingold E.A.; Grouse L.H.; Derge J.G.;  
 RA Klausner R.D.; Collins F.S.; Wagner L.; Shenmen C.M.; Schuler G.D.;  
 RA Altschul S.F.; Zeeberg B.; Buetow K.H.; Schaefer C.F.; Bhat N.K.;  
 RA Hopkins R.F.; Jordan H.; Moore T.; Max S.I.; Wang J.; Hsieh F.;  
 RA Diatchenko L.; Marusina K.; Farmer A.A.; Rubin G.M.; Hong L.;  
 RA Scapleton M.; Soares M.B.; Donald M.F.; Casavant T.L.; Scheetz T.E.;  
 RA Brownstein M.J.; Udwin T.B.; Toshiyuki S.; Carninci P.; Prange C.;  
 RA Raha S.S.; Loquellano N.A.; Peters G.J.; Abramson R.D.; Mullahy S.J.;  
 RA Bosak S.A.; McEwan P.J.; McKernan K.J.; Malek J.A.; Gunaratne P.H.;  
 RA Richards S.; Worley K.C.; Hale S.; Garcia A.M.; Gay L.J.; Huiyk S.W.;  
 RA Villalon D.K.; Muzny D.M.; Sodergren E.J.; Lu X.; Gibbs R.A.;  
 RA Fahey J.; Helton E.; Kettman M.; Madan A.; Rodriguez S.; Sanchez A.;  
 RA Whiting M.; Madan A.; Young A.C.; Shevchenko Y.; Bouffard G.G.;  
 RA Blakesley R.W.; Touchman J.W.; Green E.D.; Dickson M.C.;  
 RA Rodriguez A.C.; Grimwood J.; Schmutz J.; Myers R.M.; Butterfield Y.S.;  
 RA Krzywinski M.I.; Skalska U.; Smailus D.E.; Schnerch A.; Schein J.E.;  
 RA Jones S.J.; Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Salivary gland;  
 RA Strausberg R.;  
 RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC055451; AAH55451.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 380 AA; 43158 MW; CE1F01CA16CE5AC1 CRC64;

Query Match 34.7%; Score 67; DB 11; Length 380;  
 Best Local Similarity 55.0%; Pred. No. 3.8;  
 Matches 11; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

QY 15 EEEDDDNDEVDQAEQOKT 34  
 ||||| : : : : :  
 Db 36 EEEDDDDEDEEEKEKS 55

RESULT 13

Q9SDZ5 PRELIMINARY; PRT; 396 AA.

ID Q9SDZ5;  
 AC Q9SDZ5;  
 DT 01-MAY-2000 (TREMELrel. 13, Created)  
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)  
 DE Growth-regulating factor 1.



```

QY 15 REEDDDNEDEVDEQAQQK 33
      :||||:|||||:|::
Db 142 EDEDDDEDEDEDEEE 160

RESULT 15
000164 PRELIMINARY; PRT; 654 AA.
ID 000164
AC 000164;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Ribosomal RNA upstream binding transcription factor (Fragment).
GN UBEF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
OX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92044316; PubMed=1940801;
RA Chan E.K., Imai H., Hamel J.C., Tan E.M.;
RT "Human autoantibody to RNA polymerase I transcription factor hUBF.
RT Molecular identity of nucleolus organizer region autoantigen NOR-90
RT and ribosomal RNA transcription upstream binding factor.";
RT J. Exp. Med. 174:1239-1244(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92717570; PubMed=9126496;
RA Matera A.G., Wu W., Imai H., O'Keefe C.L., Chan E.K.;
RT "Molecular cloning of the RNA polymerase I transcription factor
RT hUBF/NOR-90 (UBTF) gene and localization to 17q21.3 by fluorescence in
RT situ hybridization and radiation hybrid mapping.";
RT Genomics 41:135-138(1997).
DR EMBL; U65487; AAC51239.1; -.
DR HSSP; P07155; 1HMF.
DR GO; GO:0003677; P:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 4.
DR SMART; SM00398; HMG; 5.
DR PROSITE; PS01118; HMG_BOX_2; 5.
FT NON_TER
FT 1
SQ SEQUENCE 654 AA; 75939 MW; 27F525850BC5FP901 CRC64;

Query Match 34.2%; Score 66; DB 4; Length 654;
Best Local Similarity 40.0%; Pred. No. 8.6;
Matches 12; Conservative 7; Mismatches 11; Indels 0; Gaps

QY 4 KGATALEINLEEDDDNEDEVDEQAQQK 33
      :||||:|||||:|::
Db 507 KSSRTTLOKSEDEDEDEDEEE 536

Search completed: March 7, 2004, 13:37:37
Job time : 42.2527 secs

```

```

Query Match          34.2%;   Score 66;   DB 4;   Length 654;
Best Local Similarity 40.0%;   Pred. NO. 8.6;
Matches 12;   Conservative 7;   Mismatches 11;   Indels 0;   Gaps 0;

QY      4   KGAAATALEINLEEDDDNDEVDQAEBOOK 33
      | | | | | | | | | | | | | | | |
DB      507 KSRRTTLOKSESEDEDEDEDEDEEE 536

Search completed: March 7, 2004, 13:37:37
Job time : 42.2527 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

un on: March 7, 2004, 13:32:36 ; Search time 17.0769 Seconds  
(without alignments)  
111.856 Million cell updates/sec

itle: US-09-643-458B-9  
erfect score: 193  
equence: 1 CKKGGAATALEINLEEDDDNEDEYDEQAEQOKTHVF 37

coring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

earched: 389414 seqs, 51625971 residues

otal number of hits satisfying chosen parameters: 389414

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : Issued Patents AA:\*

- 1: /cgm2\_6/ptodata/2/iaa/5A.COMB.pep:\*
- 2: /cgm2\_6/ptodata/2/iaa/5B.COMB.pep:\*
- 3: /cgm2\_6/ptodata/2/iaa/6A.COMB.pep:\*
- 4: /cgm2\_6/ptodata/2/iaa/6B.COMB.pep:\*
- 5: /cgm2\_6/ptodata/2/iaa/PCUTS.COMB.pep:\*
- 6: /cgm2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	167	86.5	189	3	US-08-816-346-8
2	167	86.5	189	3	US-09-335-411-8
3	167	86.5	951	3	US-08-816-346-58
4	167	86.5	951	3	US-09-335-411-58
5	167	86.5	952	2	US-08-788-674-5
6	167	86.5	952	3	US-08-816-346-4
7	167	86.5	952	3	US-09-335-411-4
8	162	83.9	45	3	US-08-816-346-12
9	162	83.9	45	3	US-09-335-411-12
10	66	34.2	764	4	US-09-370-838-67
11	64	33.2	51	3	US-08-816-346-10
12	64	33.2	51	3	US-09-335-411-10
13	64	33.2	201	3	US-08-816-346-6
14	64	33.2	201	3	US-09-335-411-6
15	64	33.2	478	3	US-09-093-448-2
16	64	33.2	478	4	US-09-813-555-2
17	64	33.2	478	4	US-09-523-263B-17
18	64	33.2	967	3	US-08-816-346-56
19	64	33.2	967	3	US-09-335-411-56
20	64	33.2	968	3	US-08-816-346-2
21	64	33.2	968	3	US-09-335-411-2
22	63	32.6	478	3	US-09-093-448-1
23	63	32.6	478	4	US-09-813-555-1
24	63	32.6	478	4	US-09-523-263B-16
25	62	32.1	200	4	US-09-702-705-324
26	62	32.1	200	4	US-09-702-705-789
27	62	32.1	200	4	US-09-736-457-324

28	62	32.1	200	4	US-09-736-457-789	Sequence 789, App
29	62	32.1	200	4	US-09-614-124B-324	Sequence 324, App
30	62	32.1	200	4	US-09-614-124B-789	Sequence 789, App
31	62	32.1	200	4	US-09-614-124B-324	Sequence 324, App
32	62	32.1	200	4	US-09-614-124B-789	Sequence 789, App
33	62	32.1	200	4	US-09-589-184-324	Sequence 324, App
34	62	32.1	200	4	US-09-589-184-789	Sequence 789, App
35	62	32.1	207	4	US-09-702-705-1667	Sequence 1667, App
36	62	32.1	207	4	US-09-736-457-1667	Sequence 1667, App
37	62	32.1	207	4	US-09-614-124B-1667	Sequence 1667, App
38	62	32.1	207	4	US-09-614-124B-1667	Sequence 1667, App
39	61	31.6	279	4	US-09-699-266A-7	Sequence 7, Appli
40	60.5	31.3	1719	2	US-08-459-568-4	Sequence 4, Appli
41	60.5	31.3	1719	2	US-08-399-411-4	Sequence 4, Appli
42	60.5	31.3	1719	3	US-08-516-859A-4	Sequence 4, Appli
43	60.5	31.3	1719	4	US-09-586-472-4	Sequence 4, Appli
44	60.5	31.3	1719	4	US-09-528-706-4	Sequence 4, Appli
45	60.5	31.3	1985	4	US-09-495-714C-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-08-816-346-8  
; Sequence 8, Application US/08816346  
; Patent No. 6127525  
; GENERAL INFORMATION:  
; APPLICANT: Crystal, Ronald G.  
; APPLICANT: Falck-Pedersen, Erik  
; APPLICANT: Gall, Jason  
; APPLICANT: Koveshdi, Imre  
; APPLICANT: Wickham, Thomas J.  
; TITLE OF INVENTION: CHIMERIC ADENOVIRAL COAT PROTEIN AND  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LEYDIG VOIT & MAYER, LTD.  
; STREET: TWO PRUDENTIAL PLAZA - 4900  
; CITY: CHICAGO  
; STATE: ILLINOIS  
; COUNTRY: USA  
; ZIP: 60601-6780  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/816,346  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION: 67167  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/616-5600  
; TELEFAX: 312/616-5700  
; TELEX: 25-3533  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 189 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-816-346-8

Query Match 86.5%; Score 167; DB 3; Length 189;  
Best Local Similarity 89.2%; Pred. No. 1.5e-14;  
Matches 33; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CKKGGAATALEINLEEDDDNEDEYDEQAEQOKTHVF 37  
Db 2 CEWDEATALEINLEEDDDNEDEYDEQAEQOKTHVF 38

RESULT 2  
US-09-335-411-8  
; Sequence 8, Application US/09335411  
; Patent No. 6153435  
; GENERAL INFORMATION:  
; APPLICANT: Crystal, Ronald G.  
; APPLICANT: Falck-Pedersen, Erik  
; APPLICANT: Gall, Jason  
; APPLICANT: Kovesdi, Imre  
; APPLICANT: Wickham, Thomas J.  
; TITLE OF INVENTION: CHIMERIC ADENOVIRAL COAT PROTEIN AND  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LEYDIG, VOIT & MAYER, LTD.  
; STREET: TWO PRUDENTIAL PLAZA - 4900  
; CITY: CHICAGO  
; STATE: ILLINOIS  
; COUNTRY: USA  
; ZIP: 60601-6780  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/335,411  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/816,346  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; REFERENCE/DOCKET NUMBER: 67167  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/616-5600  
; TELEFAX: 312/616-5700  
; TELEX: 25-3533  
; INFORMATION FOR SEQ ID NO: 8:  
; LENGTH: 189 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-335-411-8  
Query Match 86.5%; Score 167; DB 3; Length 189;  
Best Local Similarity 89.2%; Pred. No. 1.5e-14;  
Matches 33; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 CKGKGATALEINLEEDDDNEDEVDEQAEQKTHVF 37  
DB 2 CEWDEAATALEINLEEDDDNEDEVDEQAEQKTHVF 38

RESULT 3  
US-08-816-346-58  
; Sequence 58, Application US/08816346  
; Patent No. 6127525  
; GENERAL INFORMATION:  
; APPLICANT: Crystal, Ronald G.  
; APPLICANT: Falck-Pedersen, Erik  
; APPLICANT: Gall, Jason  
; APPLICANT: Kovesdi, Imre  
; APPLICANT: Wickham, Thomas J.  
; TITLE OF INVENTION: CHIMERIC ADENOVIRAL COAT PROTEIN AND  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LEYDIG, VOIT & MAYER, LTD.  
; STREET: TWO PRUDENTIAL PLAZA - 4900

CITY: CHICAGO  
STATE: ILLINOIS  
COUNTRY: USA  
ZIP: 60601-6780  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/816,346  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
REFERENCE/DOCKET NUMBER: 67167  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/616-5600  
TELEFAX: 312/616-5700  
TELEX: 25-3533  
INFORMATION FOR SEQ ID NO: 58:  
LENGTH: 951 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-816-346-58  
Query Match 86.5%; Score 167; DB 3; Length 951;  
Best Local Similarity 89.2%; Pred. No. 1e-13;  
Matches 33; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 CKGKGATALEINLEEDDDNEDEVDEQAEQKTHVF 37  
DB 132 CEWDEAATALEINLEEDDDNEDEVDEQAEQKTHVF 168

RESULT 4  
US-09-335-411-58  
; Sequence 58, Application US/09335411  
; Patent No. 6153435  
; GENERAL INFORMATION:  
; APPLICANT: Crystal, Ronald G.  
; APPLICANT: Falck-Pedersen, Erik  
; APPLICANT: Gall, Jason  
; APPLICANT: Kovesdi, Imre  
; APPLICANT: Wickham, Thomas J.  
; TITLE OF INVENTION: CHIMERIC ADENOVIRAL COAT PROTEIN AND  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LEYDIG, VOIT & MAYER, LTD.  
; STREET: TWO PRUDENTIAL PLAZA - 4900  
; CITY: CHICAGO  
; STATE: ILLINOIS  
; COUNTRY: USA  
; ZIP: 60601-6780  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/335,411  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/816,346  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; REFERENCE/DOCKET NUMBER: 67167  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/616-5600  
; TELEFAX: 312/616-5700

TELEX: 25-3533  
INFORMATION FOR SEQ ID NO: 58:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 951 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
IS-09-335-411-58

Query Match 86.5%; Score 167; DB 3; Length 951;  
Best Local Similarity 89.2%; Pred. No. 1e-13;  
Matches 33; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Y 1 CKGKGATALEINLEEDDDNEDEVDEQAEQKTHVF 37  
|: |||||  
Db 132 CEWDEAATALEINLEEDDDNEDEVDEQAEQKTHVF 168

RESULT 5  
IS-08-788-674-5  
Sequence 5, Application US/08788674  
Patent No. 592315  
GENERAL INFORMATION:  
APPLICANT: ROY, Soumitra  
TITLE OF INVENTION: Adenoviruses Having Altered  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Carella, Byrne, Bain,  
ADDRESSEE: Gilfillan, Cecchi, Stewart &  
ADDRESSEE: Olstein  
STREET: 6 Becker Farm Road  
CITY: Roseland  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07068

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/788,674  
FILING DATE: 24-JAN-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Olstein, Elliot M.  
REGISTRATION NUMBER: 24,025  
REFERENCE/DOCKET NUMBER: 271010-363  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 973-994-1700  
TELEFAX: 973-994-1744  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 952 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: predicted hexon protein sequence  
NAME/KEY: for human Adenovirus 5  
IS-08-788-674-5

Query Match 86.5%; Score 167; DB 2; Length 952;  
Best Local Similarity 89.2%; Pred. No. 1e-13;  
Matches 33; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Y 1 CKGKGATALEINLEEDDDNEDEVDEQAEQKTHVF 37  
|: |||||

Db 133 CEWDEAATALEINLEEDDDNEDEVDEQAEQKTHVF 169

RESULT 6  
US-08-816-346-4  
Sequence 4, Application US/08816346  
Patent No. 6127525  
GENERAL INFORMATION:  
APPLICANT: Crystal, Ronald G.  
APPLICANT: Falck-Pedersen, Erik  
APPLICANT: Gall, Jason  
APPLICANT: Kovesdi, Imre  
APPLICANT: Wickham, Thomas J.  
TITLE OF INVENTION: CHIMERIC ADENOVIRAL COAT PROTEIN AND  
METHODS OF USING SAME  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LEYDIG, VOIT & MAYER, LTD.  
STREET: TWO PRUDENTIAL PLAZA - 4900  
CITY: CHICAGO  
STATE: ILLINOIS  
COUNTRY: USA  
ZIP: 60601-6780

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/816,346  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
REFERENCE/DOCKET NUMBER: 67167  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/616-5600  
TELEFAX: 312/616-5700  
TELEX: 25-3533

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 952 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 951,952  
OTHER INFORMATION: /note= "Xaa can be either Gln, His, or Thr"  
US-08-816-346-4

Query Match 86.5%; Score 167; DB 3; Length 952;  
Best Local Similarity 89.2%; Pred. No. 1e-13;  
Matches 33; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Y 1 CKGKGATALEINLEEDDDNEDEVDEQAEQKTHVF 37  
|: |||||

Db 133 CEWDEAATALEINLEEDDDNEDEVDEQAEQKTHVF 169

RESULT 7  
US-09-335-411-4  
Sequence 4, Application US/09335411  
Patent No. 6153435  
GENERAL INFORMATION:  
APPLICANT: Crystal, Ronald G.  
APPLICANT: Falck-Pedersen, Erik  
APPLICANT: Gall, Jason  
APPLICANT: Kovesdi, Imre  
APPLICANT: Wickham, Thomas J.  
TITLE OF INVENTION: CHIMERIC ADENOVIRAL COAT PROTEIN AND  
METHODS OF USING SAME  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:

ADDRESSER: LEYDIG, VOIT & MAYER, LTD.  
STREET: TWO PRUDENTIAL PLAZA - 4900  
CITY: CHICAGO  
STATE: ILLINOIS  
COUNTRY: USA  
ZIP: 60601-6780  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/335,411  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/816,346  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
REFERENCE/DOCKET NUMBER: 67167  
TELEPHONE: 312/616-5600  
TELEFAX: 312/616-5700  
TELEX: 25-3533  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 952 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 951,952  
OTHER INFORMATION: /note= "Xaa can be either Gln, His, or Thr"  
US-09-335-411-4

Query Match 86.5%; Score 167; DB 3; Length 952;  
Best Local Similarity 89.2%; Pred. No. 1e-13;  
Matches 33; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CKGKAATALEINLEEDDDNEDEVDQAEQKTHVF 37  
DB 133 CEWDEATALEINLEEDDDNEDEVDQAEQKTHVF 169

RESULT 8  
US-08-816-346-12  
; Sequence 12, Application US/08816346  
; Patent No. 6127525  
; GENERAL INFORMATION:  
; APPLICANT: Crystal, Ronald G.  
; APPLICANT: Falck-Pedersen, Erik  
; APPLICANT: Gall, Jason  
; APPLICANT: Kovesdi, Imre  
; APPLICANT: Wickham, Thomas J.  
; TITLE OF INVENTION: CHIMERIC ADENOVIRAL COAT PROTEIN AND  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: LEYDIG, VOIT & MAYER, LTD.  
; STREET: TWO PRUDENTIAL PLAZA - 4900  
; CITY: CHICAGO  
; STATE: ILLINOIS  
; COUNTRY: USA  
; ZIP: 60601-6780  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/816,346  
; FILING DATE:

; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; REFERENCE/DOCKET NUMBER: 67167  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/616-5600  
; TELEFAX: 312/616-5700  
; TELEX: 25-3533  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 45 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-816-346-12  
  
Query Match 83.9%; Score 162; DB 3; Length 45;  
Best Local Similarity 100.0%; Pred. No. 1.3e-14;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 6 AATALEINLEEDDDNEDEVDQAEQKTHVF 37  
DB 1 AATALEINLEEDDDNEDEVDQAEQKTHVF 32

RESULT 9  
US-09-335-411-12  
; Sequence 12, Application US/09335411  
; Patent No. 6153435  
; GENERAL INFORMATION:  
; APPLICANT: Crystal, Ronald G.  
; APPLICANT: Falck-Pedersen, Erik  
; APPLICANT: Gall, Jason  
; APPLICANT: Kovesdi, Imre  
; APPLICANT: Wickham, Thomas J.  
; TITLE OF INVENTION: CHIMERIC ADENOVIRAL COAT PROTEIN AND  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: LEYDIG, VOIT & MAYER, LTD.  
; STREET: TWO PRUDENTIAL PLAZA - 4900  
; CITY: CHICAGO  
; STATE: ILLINOIS  
; COUNTRY: USA  
; ZIP: 60601-6780  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/335,411  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/816,346  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; REFERENCE/DOCKET NUMBER: 67167  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/616-5600  
; TELEFAX: 312/616-5700  
; TELEX: 25-3533  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 45 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-09-335-411-12

Query Match 83.9%; Score 162; DB 3; Length 45;  
Best Local Similarity 100.0%; Pred. No. 1.3e-14;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2/ 6 AATALEINLEEDDDNEDEVDQAEQKTHVF 37  
 3b 1 AATALEINLEEDDDNEDEVDQAEQKTHVF 32

RESULT 10  
 US-09-370-838-67  
 Sequence 67, Application US/09370838  
 Patent No. 644425  
 GENERAL INFORMATION:  
 APPLICANT: Reed, Steven G.  
 APPLICANT: Lodes, Michael J.  
 APPLICANT: Mohamath, Roadoh  
 APPLICANT: Secrist, Heather  
 TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF  
 TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE  
 FILE REFERENCE: 210121.475C1  
 CURRENT APPLICATION NUMBER: US/09/370,838  
 CURRENT FILING DATE: 1999-08-09  
 EARLIER APPLICATION NUMBER: US 09/285,323  
 EARLIER FILING DATE: 1999-04-02  
 NUMBER OF SEQ ID NOS: 289  
 SOFTWARE: PastSeq for Windows Version 3.0  
 SEQ ID NO 67  
 LENGTH: 764  
 TYPE: PRT  
 ORGANISM: Homo sapien  
 US-09-370-838-67

Query Match 34.2%; Score 66; DB 4; Length 764;  
 Best Local Similarity 40.0%; Pred. No. 1.5;  
 Matches 12; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

2/ 4 KGAATALEINLEEDDDNEDEVDQAEQK 33  
 3b 665 KSRRTTQSKSEDEDEDEDEDEE 694

RESULT 11  
 US-08-816-346-10  
 Sequence 10, Application US/08816346  
 Patent No. 6127525  
 GENERAL INFORMATION:  
 APPLICANT: Crystal, Ronald G.  
 APPLICANT: Falck-Pedersen, Erik  
 APPLICANT: Gall, Jason  
 APPLICANT: Kovesdi, Imre  
 APPLICANT: Wickham, Thomas J.  
 TITLE OF INVENTION: CHIMERIC ADENOVIRAL COAT PROTEIN AND  
 TITLE OF INVENTION: METHODS OF USING SAME  
 NUMBER OF SEQUENCES: 60  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: LEYDIG, VOIT & MAYER, LTD.  
 STREET: TWO PRUDENTIAL PLAZA - 4900  
 CITY: CHICAGO  
 STATE: ILLINOIS  
 COUNTRY: USA  
 ZIP: 60601-6780  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/816,346  
 FILING DATE:  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 REFERENCE/DOCKET NUMBER: 67167  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312/616-5600  
 TELEFAX: 312/616-5700

TELEX: 25-3533  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 51 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-08-816-346-10

Query Match 33.2%; Score 64; DB 3; Length 51;  
 Best Local Similarity 37.5%; Pred. No. 0.11;  
 Matches 15; Conservative 9; Mismatches 4; Indels 12; Gaps 2;

QY 5 GAATALEINLEEDDDNEDEVDQAEQ-----QKTHVF 37  
 DB 5 GRAVA-----EDEDEDEDEEEEQNARDQATKTHVY 39

RESULT 12  
 US-09-335-411-10  
 Sequence 10, Application US/09335411  
 Patent No. 6153435  
 GENERAL INFORMATION:  
 APPLICANT: Crystal, Ronald G.  
 APPLICANT: Falck-Pedersen, Erik  
 APPLICANT: Gall, Jason  
 APPLICANT: Kovesdi, Imre  
 APPLICANT: Wickham, Thomas J.  
 TITLE OF INVENTION: CHIMERIC ADENOVIRAL COAT PROTEIN AND  
 TITLE OF INVENTION: METHODS OF USING SAME  
 NUMBER OF SEQUENCES: 60  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: LEYDIG, VOIT & MAYER, LTD.  
 STREET: TWO PRUDENTIAL PLAZA - 4900  
 CITY: CHICAGO  
 STATE: ILLINOIS  
 COUNTRY: USA  
 ZIP: 60601-6780  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/335,411  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION NUMBER: 08/816,346  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 REFERENCE/DOCKET NUMBER: 67167  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312/616-5600  
 TELEFAX: 312/616-5700  
 TELEX: 25-3533  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 51 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-09-335-411-10

Query Match 33.2%; Score 64; DB 3; Length 51;  
 Best Local Similarity 37.5%; Pred. No. 0.11;  
 Matches 15; Conservative 9; Mismatches 4; Indels 12; Gaps 2;

QY 5 GAATALEINLEEDDDNEDEVDQAEQ-----QKTHVF 37  
 DB 5 GRAVA-----EDEDEDEDEEEEQNARDQATKTHVY 39

RESULT 13  
US-08-816-346-6  
; Sequence 6, Application US/08816346  
; Patent No. 6127525  
; GENERAL INFORMATION:  
; APPLICANT: Crystal, Ronald G.  
; APPLICANT: Falck-Pedersen, Erik  
; APPLICANT: Gall, Jason  
; APPLICANT: Kovsed, Imre  
; APPLICANT: Wickham, Thomas J.  
; TITLE OF INVENTION: CHIMERIC ADENOVIRAL COAT PROTEIN AND  
; TITLE OF INVENTION: METHODS OF USING SAME  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LEYDIG, VOIT & MAYER, LTD.  
; STREET: TWO PRUDENTIAL PLAZA - 4900  
; CITY: CHICAGO  
; STATE: ILLINOIS  
; COUNTRY: USA  
; ZIP: 60601-6780  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/816,346  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; REFERENCE/DOCKET NUMBER: 67167  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/616-5600  
; TELEFAX: 312/616-5700  
; TELEX: 25-3533  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 201 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-816-346-6  
Query Match 33.2%; Score 64; DB 3; Length 201;  
Best Local Similarity 37.5%; Pred. No. 0.56;  
Matches 15; Conservative 9; Mismatches 4; Indels 12; Gaps 2;  
QY 5 GAATALEINLEEDDDNEVDEQAEQ-----QKTHVF 37  
Db 11 GRAVA-----EDEEEDEDEEEEQNARDQATKKTTHVY 45  
RESULT 14  
US-09-335-411-6  
; Sequence 6, Application US/09335411  
; Patent No. 6153435  
; GENERAL INFORMATION:  
; APPLICANT: Crystal, Ronald G.  
; APPLICANT: Falck-Pedersen, Erik  
; APPLICANT: Gall, Jason  
; APPLICANT: Kovsed, Imre  
; APPLICANT: Wickham, Thomas J.  
; TITLE OF INVENTION: CHIMERIC ADENOVIRAL COAT PROTEIN AND  
; TITLE OF INVENTION: METHODS OF USING SAME  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LEYDIG, VOIT & MAYER, LTD.  
; STREET: TWO PRUDENTIAL PLAZA - 4900  
; CITY: CHICAGO  
; STATE: ILLINOIS  
; COUNTRY: USA  
; ZIP: 60601-6780  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/335,411  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION NUMBER: 08/816,346  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; REFERENCE/DOCKET NUMBER: 67167  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/616-5600  
; TELEFAX: 312/616-5700  
; TELEX: 25-3533  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 201 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-335-411-6  
Query Match 33.2%; Score 64; DB 3; Length 201;  
Best Local Similarity 37.5%; Pred. No. 0.56;  
Matches 15; Conservative 9; Mismatches 4; Indels 12; Gaps 2;  
QY 5 GAATALEINLEEDDDNEVDEQAEQ-----QKTHVF 37  
Db 11 GRAVA-----EDEEEDEDEEEEQNARDQATKKTTHVY 45  
RESULT 15  
US-09-093-448-2  
; Sequence 2, Application US/09093448A  
; Patent No. 6207704  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jun O.  
; APPLICANT: Griffith, Eric C.  
; APPLICANT: Su, Zhuang  
; TITLE OF INVENTION: Type 2 Methionine Aminopeptidase (MetAP2) Inhibitors  
; TITLE OF INVENTION: and Uses Thereof  
; FILE REFERENCE: 0492611-0346  
; CURRENT APPLICATION NUMBER: US/09/093,448A  
; CURRENT FILING DATE: 1998-06-08  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 478  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
US-09-093-448-2  
Query Match 33.2%; Score 64; DB 3; Length 478;  
Best Local Similarity 45.2%; Pred. No. 1.5;  
Matches 14; Conservative 7; Mismatches 10; Indels 0; Gaps 0;  
QY 2 KKGGAATALEINLEEDDDNEVDEQAEQ 32  
Db 44 KKGGAATALEINLEEDDDNEVDEQAEQ 74  
Search completed: March 7, 2004, 13:39:18  
Job time: 17.0769 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

Run on: March 7, 2004, 13:31:21 ; Search time 20.6703 Seconds  
(without alignments)  
290.022 Million cell updates/sec

Title: US-09-643-458B-12

Effect score: 99

Sequence: 1 CKGKGILVKQNGKLESQ 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

SPTREMBL\_25.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phage.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvrius.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	70	70.7	952	12 Q80RJ2	Q80RJ2 human adeno
2	70	70.7	952	12 Q80RI5	Q80RI5 human adeno
3	70	70.7	952	12 Q80RI0	Q80RI0 human adeno
4	70	70.7	952	12 Q805J1	Q805J1 human adeno
5	52	52.5	117	8 Q9XKY4	Q9XKY4 phytomonas
6	49	49.5	313	5 Q9U8B2	Q9U8B2 hexamita sp
7	47	47.5	379	11 Q7TMA7	Q7TMA7 mus musculus
8	47	47.5	409	5 Q9U062	Q9U062 giardia lam
9	47	47.5	446	16 Q88MB5	Q88MB5 pseudomonas
10	47	47.5	473	11 Q7TSP9	Q7TSP9 mus musculus
11	47	47.5	477	11 Q7TSQ7	Q7TSQ7 mus musculus
12	47	47.5	504	11 Q7TSQ0	Q7TSQ0 mus musculus
13	47	47.5	534	11 Q7TSQ1	Q7TSQ1 mus musculus
14	46	46.5	194	10 Q8Y68	Q8Y68 arabidopsis
15	46	46.5	499	13 Q3YH11	Q3YH11 gallus gall
16	46	46.5	510	11 Q8RIE3	Q8RIE3 mus musculus

Q8IW48 homo sapien  
Q921Z4 mus musculus  
Q920L2 rattus norv  
Q921P5 mus musculus  
Q8HW3 macaca fasc  
Q493L9 arabidopsis  
Q8K2B3 mus musculus  
Q9MI97 arabidopsis  
Q9SFE1 arabidopsis  
Q9C536 arabidopsis  
Q9M2D1 arabidopsis  
Q9C739 arabidopsis  
Q9SXB2 arabidopsis  
Q54496 streptomyce  
Q8H1L4 medicago tr  
Q07453 rous sarcom  
Q03819 rous sarcom  
Q64984 rous sarcom  
Q8NM81 corynebacte  
Q8XFL5 salmonella  
Q8XQ0 escherichia  
Q8FA16 escherichia  
Q8P37 shigella fi  
Q47482 escherichia  
Q59511 pyrococcus  
Q8T075 drosophila  
Q9S8E5 drosophila  
Q9V3P6 drosophila  
Q8IGU4 drosophila

## ALIGNMENTS

## RESULT 1

Q80RJ2 ID Q80RJ2 PRELIMINARY; PRT; 952 AA.  
AC Q80RJ2;  
DT 01-JUN-2003 (TREMBLrel. 24, Created)  
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Hexon.  
OS Human adenovirus type 5.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
OX NCBI\_TaxID=28285;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KNH Ad95/5;  
RA Na B.-K., Lee J.-Y., Kim J.-H., Kim W.-J., Kang C.;  
RT "Sequence analysis of hexon protein of adenoviruses isolated in Korea.";  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF542109; AAC24084.1; -;  
DR GO; GO:0019028; C:Viral capsid; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR000736; Adeno\_hexon.  
DR Pfam; PF01065; Adeno\_hexon; 1.  
DR Pfam; PF03678; Adeno\_hexon; C: 1.  
DR ProDom; PD002815; Adeno\_hexon; 1.  
SQ SEQUENCE 952 AA; 107938 MW; E95940D564209EAB CRC64;

Query Match 70.7%; Score 70; DB 12; Length 952;

Best Local Similarity 88.2%; Pred. No. 0.023; Mismatches 2; Indels 0; Gaps 0;

Matches 15; Conservative 0;

Qy 3 GKGILVKQNGKLESQ 19

Db 245 GKGILVKQNGKLESQ 261

RESULT 2

Q80RI5 ID Q80RI5 PRELIMINARY; PRT; 952 AA.



Q80R15;  
 DT 01-JUN-2003 (TReMBLrel. 24, Created)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE Hexon.  
 OS Human adenovirus type 5.  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
 OX NCBI\_TaxID=28285;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KNH Ad99/12;  
 RA Na B.-K., Lee J.-Y., Kim J.-H., Kim W.-J., Kang C.;  
 RT "Sequence analysis of hexon protein of adenoviruses isolated in Korea."  
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF542116; AAC24091.1; -;  
 DR GO; GO:0019028; C:Viral capsid; IEA.  
 DR GO; GO:0005198; F:Structural molecule activity; IEA.  
 DR InterPro; IPR000736; Adeno\_hexon.  
 DR Pfam; PF01065; Adeno\_hexon.1.  
 DR Pfam; PF03678; Adeno\_hexon.C; 1.  
 DR ProDom; PD002815; Adeno\_hexon; 1.  
 SQ SEQUENCE 952 AA; 107915 MW; 0514B635D6CADA54 CRC64;

Query Match 70.7%; Score 70; DB 12; Length 952;  
 Best Local Similarity 88.2%; Pred. No. 0.023;  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GKGGLVVKQNGKLESQ 19  
 DB 245 GQGGLVVKQNGKLESQ 261

RESULT 3  
 Q80R10  
 ID Q80R10 PRELIMINARY; PRT; 952 AA.  
 AC Q80R10;  
 DT 01-JUN-2003 (TReMBLrel. 24, Created)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE Hexon.  
 OS Human adenovirus type 5.  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
 OX NCBI\_TaxID=28285;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KNH Ad00/5;  
 RA Na B.-K., Lee J.-Y., Kim J.-H., Kim W.-J., Kang C.;  
 RT "Sequence analysis of hexon protein of adenoviruses isolated in Korea."  
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF542121; AAC24096.1; -;  
 DR GO; GO:0019028; C:Viral capsid; IEA.  
 DR GO; GO:0005198; F:Structural molecule activity; IEA.  
 DR InterPro; IPR000736; Adeno\_hexon.  
 DR Pfam; PF01065; Adeno\_hexon.1.  
 DR Pfam; PF03678; Adeno\_hexon.C; 1.  
 DR ProDom; PD002815; Adeno\_hexon; 1.  
 SQ SEQUENCE 952 AA; 107972 MW; 35440D68DC58E71C CRC64;

Query Match 70.7%; Score 70; DB 12; Length 952;  
 Best Local Similarity 88.2%; Pred. No. 0.023;  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GKGGLVVKQNGKLESQ 19  
 DB 245 GQGGLVVKQNGKLESQ 261

RESULT 4  
 Q805J1  
 ID Q805J1 PRELIMINARY; PRT; 952 AA.  
 AC Q805J1;

DT 01-JUN-2003 (TReMBLrel. 24, Created)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE Hexon.  
 OS Human adenovirus type 5.  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
 OX NCBI\_TaxID=28285;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KNH Ad00/12, KNH Ad00/19, and KNH Ad01/1;  
 RA Na B.-K., Lee J.-Y., Kim J.-H., Kim W.-J., Kang C.;  
 RT "Sequence analysis of hexon protein of adenoviruses isolated in Korea."  
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF542124; AAC24099.1; -;  
 DR EMBL; AF542128; AAC24103.1; -;  
 DR EMBL; AF542130; AAC24105.1; -;  
 DR GO; GO:0019028; C:Viral capsid; IEA.  
 DR GO; GO:0005198; F:Structural molecule activity; IEA.  
 DR InterPro; IPR000736; Adeno\_hexon.  
 DR Pfam; PF01065; Adeno\_hexon.1.  
 DR Pfam; PF03678; Adeno\_hexon.C; 1.  
 DR ProDom; PD002815; Adeno\_hexon; 1.  
 SQ SEQUENCE 952 AA; 107914 MW; C92738D3C0383802 CRC64;

Query Match 70.7%; Score 70; DB 12; Length 952;  
 Best Local Similarity 88.2%; Pred. No. 0.023;  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GKGGLVVKQNGKLESQ 19  
 DB 245 GQGGLVVKQNGKLESQ 261

RESULT 5  
 Q9XKY4  
 ID Q9XKY4 PRELIMINARY; PRT; 117 AA.  
 AC Q9XKY4;  
 DT 01-NOV-1999 (TReMBLrel. 12, Created)  
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
 DE MURF5 (Fragment).  
 GN MURF5.  
 OS Phytomonas serpens.  
 OG Mitochondrion.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Phytomonas.  
 OX NCBI\_TaxID=5707;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IG;  
 RX MEDLINE=98327209; PubMed=9662707;  
 RA Maslov D.A., Hollar L., Haghighat P., Nawathean P.;  
 RT "Demonstration of mRNA editing and localization of guide RNA genes in kinetoplast-mitochondria of the plant trypanosomatid Phytomonas serpens."  
 RL Mol. Biochem. Parasitol. 99:207-221(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IG;  
 RX MEDLINE=99270310; PubMed=10340485;  
 RA Maslov D.A., Nawathean P., Scheel J.;  
 RT "Partial kinetoplast-mitochondrial gene organization and expression in the respiratory deficient plant trypanosomatid Phytomonas serpens."  
 RL Mol. Biochem. Parasitol. 99:207-221(1999).  
 DR EMBL; AF079867; AAD28357.1; -;  
 DR GO; GO:0005739; C:mitochondrion; IEA.  
 KW Mitochondrion.  
 FT NON TER 117  
 SQ SEQUENCE 117 AA; 13728 MW; 65C375B260C999EE CRC64;

Query Match 52.5%; Score 52; DB 8; Length 117;  
 Best Local Similarity 50.0%; Pred. No. 1.9;  
 Matches 8; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 2 KGKGGILVKQONGKLE 17  
 DB 94 RRGGVGVGKEEGKIE 109

RESULT 6  
 Q9UB2 PRELIMINARY; PRT; 313 AA.  
 ID Q9UB2  
 AC Q9UB2  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Carbamate kinase (EC 2.7.2.2).  
 GN CBK  
 OS Hexamita sp.  
 OC Eukaryota, Diplomonadida; Hexamitidae; Hexamitinae; Hexamita.  
 OX NCBI\_TaxID=28003;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20454785; PubMed=11001147;  
 RA Dimopoulos M., Bagnara A.S., Edwards M.R.;  
 RT "Characterisation and sequence analysis of a carbamate kinase gene  
 from the diplomonad Hexamita inflata.";  
 RL J. Eukaryot. Microbiol. 47:499-503(2000).  
 DR EMBL; AF107491; AAF08984.1; -;  
 DR HSSP; P95474; 1E19.  
 DR GO; GO:000804; P:carbamate kinase activity; IEA.  
 DR GO; GO:0004349; P:glutamate 5-kinase activity; IEA.  
 DR GO; GO:0016740; P:transferase activity; IEA.  
 DR GO; GO:0008652; P:amino acid biosynthesis; IEA.  
 DR GO; GO:0006525; P:arginine metabolism; IEA.  
 DR GO; GO:0006561; P:proline biosynthesis; IEA.  
 DR InterPro; IPR001048; Aa kinase.  
 DR InterPro; IPR003964; Bac carb kinase.  
 DR InterPro; IPR01057; Glu-kinase.  
 DR Pfam; PF00696; aakinas; 1.  
 DR PRINTS; PR01469; CARBMTKINASE.  
 DR PRINTS; PR00474; GLU5KINASE.  
 DR TIGRFAMs; TIGR00746; arcc; 1.  
 KW Kinase; Transferase.  
 SQ SEQUENCE 313 AA; 33739 MW; 52D3AB67B7E1927B CRC64;

Query Match 49.5%; Score 49; DB 5; Length 313;  
 Best Local Similarity 58.8%; Pred. No. 16;  
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 CKGGGILVKQONGKLE 17  
 DB 190 CTNGGGIPVIEKNGKLE 206

RESULT 7  
 Q7TMA7 PRELIMINARY; PRT; 379 AA.  
 ID Q7TMA7  
 AC Q7TMA7  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Mannose receptor-like isoform 3 (Mannose receptor-like isoform 5).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
 RA Jang J.S. Sr., Ackerman S.L.;  
 RT "Alternatively spliced isoform 3 of mannose receptor precursor-like gene.";  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
 RA Jang J.S. Sr., Ackerman S.L.;  
 RT "Alternatively spliced isoform 5 of mannose receptor precursor-like gene.";  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY223870; AAP22984.1; -;  
 DR EMBL; AY223872; AAP22986.1; -;  
 KW Receptor.  
 SQ SEQUENCE 379 AA; 42164 MW; 98D94076795F85D5 CRC64;

Query Match 47.5%; Score 47; DB 11; Length 379;  
 Best Local Similarity 41.2%; Pred. No. 40;  
 Matches 7; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 1 CKGGGILVKQONGKLE 17  
 DB 260 CQKGGVLAQIESQVQ 276

RESULT 8  
 Q9U062 PRELIMINARY; PRT; 409 AA.  
 ID Q9U062  
 AC Q9U062  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE DnaJ homolog  
 OS Giardia lamblia (Giardia intestinalis).  
 OC Eukaryota, Diplomonadida; Hexamitidae; Giardiae; Giardia.  
 OX NCBI\_TaxID=5741;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-WB;  
 RA Dan M., Wang C.-C.;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF128225; AAD51092.1; -;  
 DR HSSP; P08622; 1BQZ.  
 DR GO; GO:0003754; F:chaperone activity; IEA.  
 DR InterPro; IPR002939; DnaJ\_C.  
 DR InterPro; IPR001305; DnaJ\_CXKXGKXG.  
 DR InterPro; IPR001623; DnaJ\_N.  
 DR InterPro; IPR008971; HSP40\_dnaJ\_pep.  
 DR InterPro; IPR003095; Hsp\_dnaJ.  
 DR Pfam; PF00226; DnaJ; 1.  
 DR Pfam; PF01556; DnaJ\_C; 1.  
 DR Pfam; PF00684; DnaJ\_CXKXGKXG; 1.  
 DR PRINTS; PR00625; DnaJPROTEIN.  
 DR SMART; SM00271; DnaJ; 1.  
 DR PROSITE; PS00636; DnaJ\_1; 1.  
 DR PROSITE; PS50076; DnaJ\_2; 1.  
 SQ SEQUENCE 409 AA; 45110 MW; 98EE5AD3EBE678D CRC64;

Query Match 47.5%; Score 47; DB 5; Length 409;  
 Best Local Similarity 42.1%; Pred. No. 44;  
 Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 CKGGGILVKQONGKLE 19  
 DB 152 CKGGGVVITQMGPMITQ 170

RESULT 9  
 Q8MB5 PRELIMINARY; PRT; 446 AA.  
 ID Q8MB5  
 AC Q8MB5  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Conserved hypothetical protein.  
 GN pp1659.  
 OS Pseudomonas putida (strain KT240).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: March 7, 2004, 13:28:01 ; Search time 4.8022 Seconds  
(without alignments)  
206.017 Million cell updates/sec

title: US-09-643-458b-12

effect score: 99  
sequence: 1 CKGKGILVKQNGKLEQ 19

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atbase : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

result No.	Score	Match	Length	ID	Description
1	70	70.7	951	1 HEX_ADE05	P04133 human adeno
2	51	51.5	224	1 R33_CHLMU	Q9pjm0 chlamydia m
3	51	51.5	224	1 R33_CHLTR	Q84527 chlamydia t
4	49	49.5	223	1 R33_CHLVC	Q824p5 chlamydia t
5	48	48.5	223	1 R33_CHLPN	Q927r3 chlamydia p
6	47	47.5	319	1 K6FF_LACDE	P80019 lactobacill
7	46	46.5	664	1 DHSA_HUMAN	P31040 homo sapien
8	46	46.5	665	1 DHSA_BOVIN	P31039 bos taurus
9	45	45.5	603	1 ENV_RVVP	P03396 rous sarcom
10	44.5	44.9	601	1 FRDA_ECOLI	P00363 escherichia
11	44	44.4	71	1 RL31_THEMA	O54311 thermotoga
12	44	44.4	175	1 YF67_METUA	Q58962 methanococc
13	44	44.4	851	1 NUD1_YEAST	P23336 saccharomyc
14	43	43.4	598	1 FRDA_PROVU	P20922 proteus vul
15	43	43.4	599	1 FRDA_HAEIN	P44894 haemophilus
16	43	43.4	1125	1 MPD_BORBU	O51568 borrelia bu
17	42	42.4	131	1 PROF_PRUVI	Q9xf39 prunus aviu
18	42	42.4	1418	1 SWC4_YEAST	Q12267 saccharomyc
19	41.5	41.9	118	1 RL33_ORYSA	P56724 oryza sativ
20	41.5	41.9	266	1 XKDC_BACSU	P39782 bacillus su
21	41.5	41.9	504	1 PUR9_FUSUN	Q9rev6 bifunctio
22	41	41.4	133	1 PR03_TOBAC	Q9et98 nicotiana t
23	41	41.4	134	1 PR01_TOBAC	P41372 nicotiana t
24	41	41.4	134	1 PR02_TOBAC	Q9et99 nicotiana t
25	41	41.4	323	1 MENC_YERPE	P58487 yersinia pe
26	41	41.4	594	1 UF71_DEIRA	Q9ru24 deinococcu
27	41	41.4	880	1 GRNA_HAEIN	P43700 haemophilus
28	40.5	40.9	947	1 CDAA_HUMAN	Q9un74 homo sapien
29	40	40.4	131	1 PR01_RICCO	O82572 ricinus com
30	40	40.4	131	1 PR02_HEVER	Q9etb6 hevea bras
31	40	40.4	189	1 GPI_PYRAB	Q9uxw3 pyrococcus
32	40	40.4	189	1 GPI_PYRHO	O59618 pyrococcus
33	40	40.4	246	1 YAFK_ECOLI	Q47148 escherichia

## ALIGNMENTS

### RESULT 1

HEX\_ADE05  
ID HEX\_ADE05 STANDARD; PRT; 951 AA.  
AC P04133;  
DT 01-NOV-1986 (Rel. 03, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DB Hexon protein (late protein 2).  
GN PII.  
OS Human adenovirus type 5.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
OX NCBI\_TaxID=28285;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84212465; PubMed=6202684;  
RA Kinloch R., MacKay N., Mautner V.;  
RT "Adenovirus hexon. Sequence comparison of subgroup C serotypes 2 and 5."  
RL J. Biol. Chem. 259:6431-6436(1984).  
RN [2]  
RP COMPLETE GENOME.  
RX MEDLINE=92087470; PubMed=1727603;  
RA Chroboczek J., Bieber F., Jacrot B.;  
RT "The sequence of the genome of adenovirus type 5 and its comparison with the genome of adenovirus type 2."  
RL Virology 186:280-285(1992).  
CC -!- FUNCTION: This protein is one of the structural proteins in the viral coat and is synthesized during late infection.  
CC -!- SUBUNIT: Homotrimer (By similarity).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; J01966; AAD15293.1; -  
CC EMBL; M73260; -; NOT ANNOTATED\_CDS.  
CC EMBL; X02997; CAA26753.1; -  
CC PIR; A03849; HXADS.  
CC PDB; 1P30; 11-NOV-03.  
CC InterPro; IPR000736; Adeno\_hexon.  
CC Pfam; PF01065; Adeno\_hexon; 1.  
CC Pfam; PF03678; Adeno\_hexon; 1.  
CC ProDom; PD002815; Adeno\_hexon; 1.  
CC Coat protein; Hexon protein; Late protein; 3D-structure.  
CC INIT\_MET 0 0 BY SIMILARITY.  
CC DOMAIN 133 158 ASP/GLU-RICH (ACIDIC).  
CC SQ SEQUENCE 951 AA; 107875 MW; EF0F9A24961B117F CRC64;

Query Match 70.7%; Score 70; DB 1; Length 951;  
Best Local Similarity 88.2%; Pred. No. 0.0057;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Q12829 homo sapien  
Q8vhp8 mus musculus  
Q9sel7 arabidopsis  
Q31540 bacillus su  
Q19753 caenorhabdi  
P45128 haemophilus  
P15120 gallus gall  
Q92990 rickettsia  
Q42449 arabidopsis  
P49231 phaseolus v  
Q9snw6 lilium long  
Q9snw5 lilium long

QY 3 CKGGILVKQNGKLEQ 19  
DB 244 GGGILVKQNGKLEQ 260

## RESULT 2

```
RS3_CHLMU          STANDARD;          PRT;   224 AA.
AC Q9PJMO;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S3.
GN RPSC OR TC0809.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Nigg;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae A39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA
CC in the 70S ribosome, positioning it for translation (By
CC similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex
CC with proteins S10 and S14 (By similarity).
CC -!- SIMILARITY: Belongs to the S3p family of ribosomal proteins.
CC -!- SIMILARITY: Contains 1 KH type-2 domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; AB002347; AAF39612.1; -.
DR PIR; A81665; A81665.
DR TIGR; TC0809; -.
DR HAMAP; MF_01309; -.
DR InterPro; IPR004087; KH dom.
DR InterPro; IPR009019; KH prot.
DR InterPro; IPR004044; KH TYPE 2.
DR InterPro; IPR003351; Ribosomal_S3_C.
DR InterPro; IPR008282; Ribosomal_S3_N.
DR InterPro; IPR005704; S3_bact.
DR Pfam; PF00013; KH; 1.
DR Pfam; PF00189; Ribosomal_S3_C; 1.
DR Pfam; PF00417; Ribosomal_S3_N; 1.
DR SMART; SM00322; KH; 1.
DR TIGRPFAMs; TIGR01009; rpsC_bact; 1.
DR PROSITE; PS50823; KH TYPE 2; 1.
DR PROSITE; PS00548; RIBOSOMAL_S3; 1.
DR Ribosomal protein; RNA-binding; Complete proteome.
FT DOMAIN 39 107 KH TYPE-2.
SQ SEQUENCE 224 AA; 24371 MW; 8B7658B682640784 CRC64;
```

Query Match 51.5%; Score 51; DB 1; Length 224;  
Best Local Similarity 47.1%; Pred. No. 1.2;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 CKGGILVKQNGKLE 17  
DB 49 CGAAGFVVKMSGKIE 65

## RESULT 3

```
RS3_CHLTR          STANDARD;          PRT;   224 AA.
AC O84E27;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S3.
GN RPSC OR RS3 OR CT522.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RX MEDLINE=9900809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA
CC in the 70S ribosome, positioning it for translation (By
CC similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex
CC with proteins S10 and S14 (By similarity).
CC -!- SIMILARITY: Belongs to the S3p family of ribosomal proteins.
CC -!- SIMILARITY: Contains 1 KH type-2 domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; AB001323; AAC68123.1; -.
DR PIR; H71506; H71506.
DR HAMAP; MF_01309; -.
DR InterPro; IPR004087; KH dom.
DR InterPro; IPR009019; KH prot.
DR InterPro; IPR004044; KH TYPE 2.
DR InterPro; IPR001351; Ribosomal_S3_C.
DR InterPro; IPR008282; Ribosomal_S3_N.
DR InterPro; IPR005704; S3_bact.
DR Pfam; PF00013; KH; 1.
DR Pfam; PF00189; Ribosomal_S3_C; 1.
DR Pfam; PF00417; Ribosomal_S3_N; 1.
DR SMART; SM00322; KH; 1.
DR TIGRPFAMs; TIGR01009; rpsC_bact; 1.
DR PROSITE; PS50823; KH TYPE 2; 1.
DR PROSITE; PS00548; RIBOSOMAL_S3; 1.
DR Ribosomal protein; RNA-binding; Complete proteome.
FT DOMAIN 39 107 KH TYPE-2.
SQ SEQUENCE 224 AA; 24415 MW; 1F12DC6AEDB8E1C CRC64;
```

Query Match 51.5%; Score 51; DB 1; Length 224;  
Best Local Similarity 47.1%; Pred. No. 1.2;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 CKGGILVKQNGKLE 17  
DB 49 CGAAGFVVKMSGKIE 65

## RESULT 4

RS3\_CHLCV STANDARD; PRT; 223 AA.
ID RS3\_CHLCV
AC Q824P5;

10-OCT-2003 (Rel. 42, Created)  
 10-OCT-2003 (Rel. 42, Last sequence update)  
 10-OCT-2003 (Rel. 42, Last annotation update)  
 30S ribosomal protein S3.  
 NPSC OR CCA00099.  
 NCBI\_TaxID=83557;  
 Chlamydia; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales.  
 NCBI\_TaxID=83557;  
 SEQUENCE FROM N.A.  
 STRAIN=GPIC;  
 MEDLINE=22569155; PubMed=12682364;  
 Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,  
 Heidelberg J., Holtzapple E., Khouri H., Federova N.B., Carty H.A.,  
 Unayem L.A., Haft D.H., Peterson J., Beanan M.J., White O.,  
 Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bayoill P.M.,  
 Fraser C.M.;  
 "Genome sequence of Chlamydia caviae (Chlamydia psittaci GPIC):  
 examining the role of niche-specific genes in the evolution of the  
 Chlamydiales";  
 Nucleic Acids Res. 31:2134-2147(2003).  
 -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA  
 in the 70S ribosome, positioning it for translation (By  
 similarity).  
 -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex  
 with proteins S10 and S14 (By similarity).  
 -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.  
 -!- SIMILARITY: Contains 1 KH type-2 domain.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 EMBL; AEO16994; AAP04851.1; -;  
 TIGR; CCA00099; -;  
 HAMAP; MF\_01309; -; 1.  
 InterPro; IPR004087; KH dom.  
 InterPro; IPR009019; KH\_prok.  
 InterPro; IPR004044; KH\_TYPE\_2.  
 InterPro; IPR001351; Ribosomal\_S3\_C.  
 InterPro; IPR008282; Ribosomal\_S3\_N.  
 InterPro; IPR005704; S3\_bact.  
 Pfam; PF00013; KH; 1.  
 Pfam; PF00189; Ribosomal\_S3\_C; 1.  
 Pfam; PF00417; Ribosomal\_S3\_N; 1.  
 SMART; SM00322; KH; 1.  
 TIGRFAMs; TIGR01009; rpsc\_bact; 1.  
 PROSITE; PS00823; KH\_TYPE\_2; 1.  
 PROSITE; PS00548; RIBOSOMAL\_S3; 1.  
 Ribosomal protein; RNA-binding; Complete proteome.  
 DOMAIN 39 117 KH type-2.  
 SEQUENCE 223 AA; 24620 MW; 334316FCF859FAE CRC64;  
 Query Match 49.5%; Score 49; DB 1; Length 223;  
 Best Local Similarity 41.2%; Pred. No. 2.4;  
 Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;  
 1 CKGKGKGVILKQNGKLE 17  
 49 CQGAAGFIVRMSSGKLE 65  
 RESULT 5  
 S3\_CHLNP  
 D RS3\_CHLNP STANDARD; PRT; 223 AA.  
 C Q927R3; Q9JRT5;  
 30-MAY-2000 (Rel. 39, Created)  
 16-OCT-2001 (Rel. 40, Last sequence update)  
 10-OCT-2003 (Rel. 42, Last annotation update)

30S ribosomal protein S3.  
 NPSC OR RS3 OR CPN0641 OR CP0106 OR CPB0667.  
 Chlamydia pneumoniae (Chlamydia pneumoniae).  
 Bacteria; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales.  
 NCBI\_TaxID=83558;  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN=CWL029;  
 MEDLINE=99206606; PubMed=10192388;  
 Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 "Comparative Genomes of Chlamydia pneumoniae and C. trachomatis";  
 Nat. Genet. 21:385-389(1999).  
 [2]  
 SEQUENCE FROM N.A.  
 STRAIN=AR39;  
 MEDLINE=2015255; PubMed=10684935;  
 Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 Gwinn M., Nelson W., deBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 Eisen J., Fraser C.M.;  
 "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 pneumoniae AR39";  
 Nucleic Acids Res. 28:1397-1406(2000).  
 [3]  
 SEQUENCE FROM N.A.  
 STRAIN=J138;  
 MEDLINE=20330349; PubMed=10871362;  
 Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 from Japan and CWL029 from USA";  
 Nucleic Acids Res. 28:2311-2314(2000).  
 [4]  
 SEQUENCE FROM N.A.  
 STRAIN=TW-183;  
 Geng M.M., Schumacher A., Muehldorfer I., Bensch K.W., Schaefer K.P.,  
 Schneider S., Pohl T., Essig A., Marre R., Melchers K.;  
 "The genome sequence of Chlamydia pneumoniae TW183 and comparison with  
 other Chlamydia strains based on whole genome sequence analysis";  
 Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 -!- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA  
 in the 70S ribosome, positioning it for translation (By  
 similarity).  
 -!- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex  
 with proteins S10 and S14 (By similarity).  
 -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.  
 -!- SIMILARITY: Contains 1 KH type-2 domain.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 EMBL; AEO01647; AAD18780.1; -;  
 EMBL; AEO02173; AAF37989.1; -;  
 EMBL; AP002547; BAA98848.1; -;  
 EMBL; AEO17159; AAP98596.1; -;  
 PIR; C72055; C72055.  
 PIR; F81611; F81611.  
 PIR; F86570; F86570.  
 TIGR; CP0106; -;  
 HAMAP; MF\_01309; -; 1.  
 InterPro; IPR004087; KH dom.  
 InterPro; IPR009019; KH\_prok.  
 InterPro; IPR004044; KH\_TYPE\_2.  
 InterPro; IPR001351; Ribosomal\_S3\_C.  
 InterPro; IPR008282; Ribosomal\_S3\_N.  
 InterPro; IPR005704; S3\_bact.

DR Pfam; PF00013; KH; 1.  
DR Pfam; PF00189; Ribosomal\_S3\_C; 1.  
DR Pfam; PF00417; Ribosomal\_S3\_N; 1.  
DR SMART; SM00322; KH; 1.  
DR TIGRFAMs; TIGR01009; rpsC\_bact; 1.  
DR PROSITE; PS00823; KH TYPE 2; 1.  
DR PROSITE; PS00548; RIBOSOMAL\_S3; 1.  
KW Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.  
FT DOMAIN 39 107 KH TYPE-2.  
FT CONFLICT 17 107 K -> R (IN REF. 1).  
SQ SEQUENCE 223 AA; 24584 MW; F3B42DD14E741DAC CRC64;  
  
Query Match 48.5%; Score 48; DB 1; Length 223;  
Best Local Similarity 41.2%; Pred. No. 3.4; Indels 0; Gaps 0;  
Matches 7; Conservative 6; Mismatches 4;  
  
QY 1 CKGKGILVKQONGKLE 17  
DB 49 CGAAGFVVRMSGKIE 65  
  
RESULT 6  
K6PF LACDE STANDARD; PRT; 319 AA.  
AC P80019;  
DT 01-MAY-1991 (Rel. 18, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 6-phosphofructokinase (EC 2.7.1.1) (Phosphofructokinase)  
DE (Phosphohexokinase).  
GN PFKA.  
OS Lactobacillus delbrueckii (subsp. bulgaricus).  
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
OC Lactobacillus.  
OX NCBI\_TaxID=1585;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93374827; PubMed=8366023;  
RA Branny P., de la Torre F., Garel J.R.;  
RT "Cloning, sequencing, and expression in Escherichia coli of the gene  
RT coding for phosphofructokinase in Lactobacillus bulgaricus";  
RL J. Bacteriol. 175:5344-5349(1993).  
RN [2]  
RP SEQUENCE OF 1-38.  
RC STRAIN=B107;  
RX MEDLINE=91266952; PubMed=1828763;  
RA le Bras G., Deville-Bonne D., Garel J.R.;  
RT "Purification and properties of the phosphofructokinase from  
RT Lactobacillus bulgaricus. A non-allosteric analog of the enzyme from  
RT Escherichia coli";  
RL Eur. J. Biochem. 198:693-697(1991).  
CC -!- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-fructose 1,6-bisphosphate.  
CC -!- ENZYME REGULATION: Activated by ADP and inhibited by phosphoenolpyruvate.  
CC -!- PATHWAY: Key control step of glycolysis.  
CC -!- SUBUNIT: Homotetramer.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: Belongs to the phosphofructokinase family.  
  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X71403; CAA50526.1; -;  
DR PIR; A48663; S35928.  
DR HSSP; P00512; 3PFK.  
DR HAMAP; MF00339; -;  
DR InterPro; IPR000023; Ppfructokinase.

DR Pfam; PF00365; PFK; 1.  
DR PRINTS; PR00476; PFRCTKINASE.  
DR ProDom; PD000707; Pfrctkinase; 1.  
DR PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 1.  
KW Kinase; Transferase; Glycolysis.  
SQ SEQUENCE 319 AA; 34008 MW; 555C5FE1D522BC9C CRC64;  
  
Query Match 47.5%; Score 47; DB 1; Length 319;  
Best Local Similarity 52.9%; Pred. No. 6.7;  
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
  
QY 2 KGKGILVKQONGKLES 18  
DB 277 EKGGLAVGIENGKVTYS 293  
  
RESULT 7  
DHSA HUMAN STANDARD; PRT; 664 AA.  
ID DHSA\_HUMAN  
AC P31040; Q16395; Q9UMY5;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Succinate dehydrogenase [ubiquinone] flavoprotein subunit,  
DE mitochondrial precursor (EC 1.3.5.1) (Pp) (Flavoprotein subunit of  
DE complex II).  
GN SDHA OR SDH2 OR SDHF.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=95096020; PubMed=7798181;  
RA Hirawake H., Wang H., Kuramochi T., Kojima S., Kita K.;  
RT "Human complex II (succinate-ubiquinone oxidoreductase): cDNA cloning  
RT of the flavoprotein (Pp) subunit of liver mitochondria";  
RL J. Biochem. 116:221-227(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Heart;  
RX MEDLINE=94190953; PubMed=8142412;  
RA Morris A.A.M., Farnsworth L., Ackrell B.A.C., Turnbull D.M.,  
RA Birch-Machin M.A.;  
RT "The cDNA sequence of the flavoprotein subunit of human heart  
RT succinate dehydrogenase";  
RL Biochim. Biophys. Acta 1185:125-128(1994).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC Farfait B., Chretien D., Rotig A., Munnich A., Rustin P.;  
RT "Compound heterozygous mutations in the flavoprotein gene of the  
RT respiratory chain complex II in a patient with Leigh's syndrome";  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Colon;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg K.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner K.H., Scheiner C.F., Bhat N.K.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Sapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Hale S.C., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,



A Butterfield Y.S.N., Krzyzinski M.I., Skalska U., Smalhus D.E.,  
A Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
T "Generation and initial analysis of more than 15,000 full-length  
T human and mouse cDNA sequences.";  
L Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
N [5]  
N PRELIMINARY PARTIAL SEQUENCE FROM N.A.  
C TISSUE-Placenta;  
A Malcovati M., Marchetti L., Zanelli E., Tencchini M.L.;  
C "Cloning of the flavoprotein subunit of human succinate  
T dehydrogenase.";  
T (in) Curti B., Ronchi S., Zanetti G. (eds.);  
L Flavins and flavoproteins 1990, pp.727-730, Walter de Gruyter,  
L Berlin (1991).  
N [6]  
N SEQUENCE OF 546-562 FROM N.A., AND VARIANT LS TRP-554.  
P MEDLINE=96024649; PubMed=7550341;  
X A Bourgeron T., Rustin P., Chretien D., Birch-Machin M.A., Bourgeois M.,  
A Viegas-Pequignot E., Munnich A., Roedig A.;  
T "Mutation of a nuclear succinate dehydrogenase gene results in  
T mitochondrial respiratory chain deficiency.";  
N Nat. Genet. 11:144-149(1995).  
N [7]  
N VARIANT LS VAL-524.  
P MEDLINE=20208355; PubMed=10746566;  
X A Farfait B., Chretien D., Rotig A., Marsac C., Munnich A., Rustin P.;  
A "Compound heterozygous mutations in the flavoprotein gene of the  
T respiratory chain complex II in a patient with Leigh syndrome.";  
T Hum. Genet. 106:236-243(2000).  
N [8]  
N VARIANT LS GLU-555.  
P MEDLINE=22678447; PubMed=12794685;  
X A Van Coster R., Seneca S., Smet J., Van Hecke R., Gerlo E.,  
A Devreese B., Van Beeumen J., Leroy J.G., De Meirleir L., Lissens W.;  
T "Homozygous Gly555Glu mutation in the nuclear-encoded 70 kDa  
T flavoprotein gene causes instability of the respiratory chain complex  
T II.";  
L Am. J. Med. Genet. 120:13-18(2003).  
C -!- CATALYTIC ACTIVITY: Succinate + ubiquinone = fumarate + ubiquinol.  
C -!- COFACTOR: FAD.  
C -!- PATHWAY: Tricarboxylic acid cycle.  
C -!- SUBUNIT: Composed of a 27 kDa iron protein (IP), a 70 kDa  
C flavoprotein (FP) and a cytochrome B composed of two integral  
C membrane proteins.  
C -!- SUBCELLULAR LOCATION: Mitochondrial inner membrane.  
C -!- DISEASE: Defects in SDHA are a cause of Leigh syndrome (LS)  
C [MIM:256000]. LS is a severe neurological disorder characterized  
C by bilaterally symmetrical necrotic lesions in subcortical brain  
C regions.  
C -!- MISCELLANEOUS: The complex, present in mitochondria, can be  
C degraded to form EC 1.3.99.1, which no longer reacts with  
C ubiquinone.  
C -!- SIMILARITY: Belongs to the FAD-dependent oxidoreductase family 2.  
C FRD/SDH subfamily.  
C -!- CAUTION: Ref.5 sequence differs extensively from that shown.  
C  
C This SWISS-PROT entry is copyright. It is produced through a collaboration  
C between the Swiss Institute of Bioinformatics and the EMBL outstation -  
C the European Bioinformatics Institute. There are no restrictions on its  
C use by non-profit institutions as long as its content is in no way  
C modified and this statement is not removed. Usage by and for commercial  
C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
C or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
C  
C -----  
C EMBL; D30648; BAA06332.1; -;  
C EMBL; L21936; AAA20683.1; -;  
C EMBL; AF171030; AAD51006.1; -;  
C EMBL; AF171017; AAD51006.1; JOINED.  
C EMBL; AF171018; AAD51006.1; JOINED.  
C EMBL; AF171019; AAD51006.1; JOINED.  
C EMBL; AF171020; AAD51006.1; JOINED.  
C EMBL; AF171021; AAD51006.1; JOINED.  
C EMBL; AF171022; AAD51006.1; JOINED.

DR EMBL; AF171023; AAD51006.1; JOINED.  
DR EMBL; AF171024; AAD51006.1; JOINED.  
DR EMBL; AF171025; AAD51006.1; JOINED.  
DR EMBL; AF171026; AAD51006.1; JOINED.  
DR EMBL; AF171027; AAD51006.1; JOINED.  
DR EMBL; AF171028; AAD51006.1; JOINED.  
DR EMBL; AF171029; AAD51006.1; JOINED.  
DR EMBL; BC001380; AAH01380.1; -;  
DR EMBL; X53943; CAA37886.1; -;  
DR EMBL; S79641; AAB35332.1; -;  
DR PIR; JX0336; JX0336.  
DR HSP; P00363; IFUM.  
DR Genew; HGNC:10680; SDHA.  
DR GK; P31040; -;  
DR MIM; 600857; -;  
DR MIM; 256000; -;  
DR GO; GO:0005739; C:mitochondrion; TAS.  
DR GO; GO:0005489; F:electron transporter activity; TAS.  
DR GO; GO:0009060; P:aerobic respiration; TAS.  
DR GO; GO:0006099; P:tricarboxylic acid cycle; TAS.  
DR InterPro; IPR001353; FAD\_bind2.  
DR InterPro; IPR001327; FAD\_DYK\_redox.  
DR InterPro; IPR003952; FRD/SDH\_FAD\_BS.  
DR InterPro; IPR001100; Pyr\_redox.  
DR Pfam; PF00890; FAD\_binding\_2; 1.  
DR Pfam; PF02910; succ\_DH\_flav\_C; 1.  
DR PRINTS; PR00368; PADPDR.  
DR PROSITE; PS00411; PNDRTASE1.  
DR PROSITE; PS00504; FRD\_SDH\_FAD\_BINDING; 1.  
KW Tricarboxylic acid cycle; Flavoprotein; FAD; Oxidoreductase;  
KW Electron transport; Mitochondrion; Transit peptide; Disease mutation;  
KW Leigh syndrome.  
FT TRANSIT 1 43 MITOCHONDRION (BY SIMILARITY).  
FT CHAIN 44 664 SUCCINATE DEHYDROGENASE [UBIQUINONE]  
FT NP\_BIND 64 78 FLAVOPROTEIN SUBUNIT  
FT BINDING 99 99 FAD (AMPHIPART) (POTENTIAL).  
FT ACT\_SITE 296 296 FAD (COVALENT) (BY SIMILARITY).  
FT ACT\_SITE 312 312 BY SIMILARITY.  
FT VARIANT 524 524 A -> V (in LS).  
FT VARIANT 554 554 /FTID=VAR\_016878.  
FT VARIANT 554 554 R -> W (in LS).  
FT VARIANT 555 555 /FTID=VAR\_002449.  
FT VARIANT 555 555 G -> E (in LS).  
FT CONFLICT 356 356 /FTID=VAR\_016879.  
FT CONFLICT 398 398 G -> D (IN REF. 3).  
FT CONFLICT 591 591 E -> D (IN REF. 3).  
FT CONFLICT 596 596 A -> T (IN REF. 3).  
FT CONFLICT 600 600 D -> G (IN REF. 3).  
FT CONFLICT 629 629 R -> Q (IN REF. 3).  
FT CONFLICT 629 629 Y -> F (IN REF. 2).  
FT CONFLICT 640 640 E -> G (IN REF. 3).  
FT CONFLICT 657 657 V -> I (IN REF. 2).  
SQ SEQUENCE 664 AA; 72691 MW; 180B664E3FFD0B34 CRC64;  
Query Match 46.5%; Score 46; DB 1; Length 664;  
Best Local Similarity 58.3%; Pred. No. 19;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 CKGKGGLVKKQ 12  
|:|:|:|:|:  
Db 311 CRGEGGILNSQ 322  
RESULT 8  
DHSA\_BOVIN STANDARD; PRT; 665 AA.  
ID DHSA\_BOVIN  
AC P31039; Q9TUY8;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Succinate dehydrogenase [ubiquinone] flavoprotein subunit,

DE mitochondrial precursor (EC 1.3.5.1) (FP) (Flavoprotein subunit of  
GN complex I);  
OS SDHA OR SDH2 OR SDHFP1.  
OC Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX TISSUE=Heart;  
RC MEDLINE=92283874; PubMed=1375942;  
RA Birch-Machin M.A., Farnsworth L., Ackrell B.A.C., Cochran B.,  
RA Jackson S., Bindoff L.A., Aitken A., Diamond A.G., Turnbull D.M.;  
RT "The sequence of the flavoprotein subunit of bovine heart succinate  
RT dehydrogenase";  
RL J. Biol. Chem. 267:11553-11558 (1992).  
RN [2]  
RP REVISIONS.  
RX MEDLINE=94190953; PubMed=8142412;  
RA Morris A.A.M., Farnsworth L., Ackrell B.A.C., Turnbull D.M.,  
RA Birch-Machin M.A.;  
RT "The cDNA sequence of the flavoprotein subunit of human heart  
RT succinate dehydrogenase";  
RL Biochim. Biophys. Acta 1185:125-128 (1994).  
RN [3]  
RP SEQUENCE OF 153-253 FROM N.A.  
RC STRAIN=Hereford x Nelore;  
RA Sonstegard T.S., Kapper S.M.;  
RT "Mapping of the SDHA locus to bovine chromosome 20";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- CATALYTIC ACTIVITY: Succinate + ubiquinone = fumarate + ubiquinol.  
CC -!- COFACTOR: FAD.  
CC -!- SUBUNIT: Tricarboxylic acid cycle.  
CC -!- SUBUNIT: Complex containing at least four different subunits: a  
CC flavoprotein, an iron-sulfur, and two hydrophobic anchor proteins.  
CC -!- SUBCELLULAR LOCATION: Mitochondrial inner membrane.  
CC -!- SIMILARITY: Belongs to the FAD-dependent oxidoreductase family 2.  
CC FRD/SDH subfamily.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M60879; AAA30758.1; -;  
CC EMBL; AF139922; RAD38150.1; -;  
CC HSP; P00363; 1FUM.  
CC InterPro; IPR003953; FAD\_bind2.  
CC InterPro; IPR001327; FAD\_Pyr\_redox.  
CC InterPro; IPR003952; FRD/SDH\_FAD\_BS.  
CC InterPro; IPR001100; Pyr\_redox.  
CC InterPro; IPR004112; Succ\_DH\_flav\_C.  
CC Pfam; PF00890; FAD\_binding\_2; 1.  
CC Pfam; PF02910; succ\_DH\_flav\_C; 1.  
CC PRINTS; P00368; FADPNE  
CC PRINTS; P00411; PNRDPTASE1.  
CC PROSITE; PS00504; FRD\_SDH\_FAD\_BINDING; 1.  
KW Tricarboxylic acid cycle; Flavoprotein; FAD; Oxidoreductase;  
KW Electron transport; Mitochondrion; Transist peptide.  
FT TRANSIT 1 44  
FT CHAIN 45 665  
FT SUCCINATE DEHYDROGENASE [UBIQUINONE]  
FT FLAVOPROTEIN SUBUNIT  
FT NP\_BIND 65 79  
FT BINDING 100 100  
FT ACT\_SITE 297 297  
FT ACT\_SITE 313 313  
FT ACT\_SITE 313 313  
SQ SEQUENCE 665 AA; 73277 MW; 41687F5A80935E0D CRC64;  
Query Match 46.5%; Score 46; DB 1; Length 665;  
OX NCBI\_TaxID=562;  
OC Enterobacteriaceae; Escherichia.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OS Escherichia coli.  
GN FRDA OR B4154.  
DE Fumarate reductase flavoprotein subunit (EC 1.3.99.1).  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DT 01-MAY-1991 (Rel. 18, Last sequence update)  
DT 21-JUL-1986 (Rel. 01, Created)  
AC P00363;  
ID FRDA\_ECOLI STANDARD; PRT; 601 AA.  
FT CHAIN 65 603  
FT CHAIN 406 603  
FT CHAIN 557 577  
FT TRANSMEM POTENTIAL.  
SQ SEQUENCE 603 AA; 65660 MW; D44EC2AA62282C94 CRC64;  
Query Match 45.5%; Score 45; DB 1; Length 603;  
Best Local Similarity 72.7%; Pred. No. 25;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CKGKGIGILVKQ 11  
DB 285 CRKGIGIWNQ 295

Best Local Similarity 58.3%; Pred. No. 19;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CKGKGIGILVKQ 12  
DB 312 CRGEGGILINSQ 323  
RESULT 9  
ENV\_RSVP  
ID ENV\_RSVP STANDARD; PRT; 603 AA.  
AC P03396;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE ENV polyprotein [Contains: Surface protein GP85; Membrane protein  
DE GP37].  
GN ENV.  
OS Rous sarcoma virus (strain Prague C).  
OS Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.  
OX NCBI\_TaxID=11888;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83155662; PubMed=6299578;  
RA Schwartz D., Tizard R., Gilbert W.;  
RT "Nucleotide sequence of Rous sarcoma virus";  
RL Cell 32:853-869 (1983).  
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC -!- MISCELLANEOUS: AS A RESULT OF BASE VARIATIONS, A DIFFERENT VERSION  
CC OF THIS SEQUENCE MAY EXIST HAVING 17-GLU, 134-SER, 158-SER,  
CC 334-THR, 383-THR, 392-VAL, 522-GLU, 541-LEU, AND 567-VAL.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; V01197; -; NOT ANNOTATED\_CDS.  
CC PIR; A03996; VCFVER.  
CC InterPro; IPR005166; Avian\_gp85.  
CC Pfam; PF03708; Avian\_gp85; 1.  
KW Coat protein; Polyprotein; Transmembrane; Signal.  
FT SIGNAL 1 64  
FT CHAIN 65 603  
FT CHAIN 406 603  
FT CHAIN 557 577  
FT TRANSMEM POTENTIAL.  
SQ SEQUENCE 603 AA; 65660 MW; D44EC2AA62282C94 CRC64;  
Query Match 45.5%; Score 45; DB 1; Length 603;  
Best Local Similarity 72.7%; Pred. No. 25;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CKGKGIGILVKQ 11  
DB 285 CRKGIGIWNQ 295



RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.,  
 RA "Evidence for lateral gene transfer between Archaea and Bacteria from  
 RT genome sequence of *Thermotoga maritima*.";  
 RL Nature 399:323-329(1999).  
 CC -!- SIMILARITY: Belongs to the L3iP family of ribosomal proteins.  
 CC Subfamily A.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL; Y11219; CAA72106.1; ALT\_INIT.  
 DR EMBL; AS001809; AAD36751.1; -.  
 DR PIR; F72222; F72222.  
 DR TIGR; TW1684; -.  
 DR HAMAP; MF 00501; -; 1.  
 DR InterPro; IPR002150; Ribosomal L31.  
 DR Pfam; PF01197; Ribosomal L31; 1.  
 DR PRINTS; PR01249; RIBOSOMALL31.  
 DR TIGRFAMs; TIGR00105; L31; 1.  
 DR PROSITE; PS01143; RIBOSOMAL\_L31; 1.  
 KW Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 71 AA; 7996 MW; 39505125D6A61256 CRC64;

Query Match 44.4%; Score 44; DB 1; Length 71;  
 Best Local Similarity 46.7%; Pred. No. 4.6;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 3 GRGGILVQKNGKLE 17  
 Db 47 GRGGVLIVDTEGRVE 61

RESULT 12  
 YP67\_METUA  
 ID YP67\_METUA STANDARD; PRT; 175 AA.  
 AC Q58962;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein MJ1567.  
 GN MJ1567.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kervlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.E., Fuhrmann J.L., Nguyen D.,  
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hrust M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, *Methanococcus*  
 RT *jannaschii*.";  
 RL Science 273:1058-1073 (1996).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL; U67597; AAB99593.1; -.  
 DR PIR; F64495; F64495.  
 DR TIGR; MJ1567; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 175 AA; 20726 MW; 20524D646C38EFD6 CRC64;

Query Match 44.4%; Score 44; DB 1; Length 175;  
 Best Local Similarity 57.1%; Pred. No. 11;  
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 4 KGGILVQKNGKLE 17  
 Db 135 KNGIIKSKNEKLE 148

RESULT 13  
 NUD1\_YEAST  
 ID NUD1\_YEAST STANDARD; PRT; 851 AA.  
 AC P32336; Q08895;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE NUD1 protein.  
 GN NUD1 OR YOR373W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Dulic V., Zanolari B., Riezman H.;  
 RT "NUD1, a cell-cycle regulated gene required for nuclear division in  
 RT Saccharomyces cerevisiae.";  
 RL Submitted (SEP-1991) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97313270; PubMed=9169874;  
 RA Dujon B., Albermann K., Aldea M., Alexandraki D., Ansoorge W.,  
 RA Arino J., Benes V., Bohn C., Bolotin-Fukuhara M., Bordonne R.,  
 RA Boyer J., Camasses A., Casanayor A., Casas C., Cheret G.,  
 RA Cziepluch C., Daignan-Fornier B., Dang D.V., de Haan M., Delius H.,  
 RA Durand P., Fairhead C.A., Feldmann H., Gailion L., Galisson F.,  
 RA Gamo F.-J., Gancedo C., Goffeau A., Goulding S.E., Grivell L.A.,  
 RA Habbig B., Hand N.J., Hani J., Hattenhorst U., Hebling U., Hofmann B.,  
 RA Hernandez Y., Herrero E., Heumann K., Hiesel R., Hilger F., Hofmann B.,  
 RA Hollenberg C.P., Hughes B., Jauniaux J.-C., Kaloogeropoulos A.,  
 RA Katsoulou C., Kordes E., Lafuente M.J., Landt O., Louis E.J.,  
 RA Maarse C., Madania A., Mannhaupt G., Marck C., Martin R.P.,  
 RA Mewes H.-W., Michaux G., Paces V., Parle-McDermott A.G., Pearson B.M.,  
 RA Perrin A., Pettersson B., Poch O., Pohl T.M., Poirey R.,  
 RA Portetelle D., Pujol A., Purnelle B., Ramezani Rad M., Rechmann S.,  
 RA Schwager C., Schweizer M., Sor F., Sterky F., Tarasov I.A.,  
 RA Teodoru C., Tettelin H., Thierry A., Tobiasch E., Izermia M.,  
 RA Uhlen M., Unselid M., Valens M., Vandenbol M., Vetter I., Vleck C.,  
 RA Voet M., Volckaert G., Voss H., Wambutt R., Wedler H., Wiemann S.,  
 RA Winsor B., Wolfe K.H., Zollner A., Zumbstein E., Kleine K.;  
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XV.";  
 RL Nature 387:98-102(1997).  
 RN [3]  
 RP IDENTIFICATION BY MASS SPECTROMETRY IN SPINDLE POLE BODY.  
 RX MEDLINE=98252849; PubMed=9585415;  
 RA Wigge P.A., Jensen O.N., Holmes S., Soues S., Mann M., Kilmartin J.V.;  
 RT "Analysis of the Saccharomyces spindle pole by matrix-assisted laser  
 RT desorption/ionization (MALDI) mass spectrometry.";  
 RL J. Cell Biol. 141:967-977(1998).  
 RN [4]  
 RP INTERACTION WITH CNM67.

XX MEDLINE-98239572; PubMed-9571234;  
TA Brachet A., Kilmartin J.V., Wach A., Philippsen P.;  
TT "Saccharomyces cerevisiae cells with defective spindle pole body  
TT outer plaques accomplish nuclear migration via half-bridge-organized  
TT microtubules."; Mol. Biol. Cell 9:977-991(1998).  
UN [5]  
UN SUBCELLULAR LOCATION.  
UN P MEDLINE-99264329; PubMed-10330408;  
XX Adams I.R., Kilmartin J.V.;  
TA "Localization of core spindle pole body (SPB) components during SPB  
TT duplication in Saccharomyces cerevisiae."; J. Cell Biol. 145:809-823(1999).  
UN [6]  
UN FUNCTION, PHOSPHORYLATION, AND INTERACTION WITH SPC72.  
UN P MEDLINE-20553227; PubMed-1101520;  
XX Grunberg U., Campbell K., Simpson C., Grindlay J., Schiebel E.;  
TA "Nudip links astral microtubule organization and the control of exit  
TT from mitosis."; EMBO J. 19:6475-6488(2000).  
UN [7]  
UN ERRATUM.  
UN P Grunberg U., Campbell K., Simpson C., Grindlay J., Schiebel E.;  
XX EMBO J. 20:305-305(2001).  
UN [8]  
UN INTERACTION WITH MPC54 AND SPO21.  
UN P MEDLINE-20359265; PubMed-10899120;  
XX Knop M., Strasser K.;  
TA "Role of the spindle pole body of yeast in mediating assembly of the  
TT prospore membrane during meiosis."; EMBO J. 19:3657-3667(2000).  
UN [9]  
UN COMPOSITION OF A SPB COMPLEX, AND INTERACTION WITH ADY3.  
UN P MEDLINE-21607787; PubMed-11742972;  
XX Moreno-Borchart A.C., Strasser K., Finkbeiner M.G., Shevchenko A.,  
TA Shevchenko A., Knop M.;  
TT "Prospore membrane formation linked to the leading edge protein (LEP)  
TT coat assembly."; EMBO J. 20:6946-6957(2001).  
UN [10]  
UN INTERACTION WITH ADY3.  
UN P MEDLINE-21969403; PubMed-11973299;  
XX Nickas M.E., Neiman A.M.;  
TA "Ady3p links spindle pole body function to spore wall synthesis in  
TT Saccharomyces cerevisiae."; Genetics 160:1439-1450(2002).  
UN [11]  
UN INTERACTION WITH ADY4.  
UN P MEDLINE-22680893; PubMed-12796288;  
XX Nickas M.E., Schwartz C., Neiman A.M.;  
TA "Ady4p and Spo7ap are components of the meiotic spindle pole body  
TT that promote growth of the prospore membrane in Saccharomyces  
TT cerevisiae."; Eukaryot. Cell 2:431-445(2003).  
UN -!- FUNCTION: Involved in astral microtubule organization by binding  
UN SPC72 to the outer plaque in a cell-cycle dependent manner.  
UN Required for the mitotic exit by facilitating the binding of TEM1  
UN to CDC15. Also involved in the pathway that organizes the shaping  
UN and sizing of the prospore membrane (PSM) during sporulation.  
UN -!- SUBUNIT: Interacts directly with MPC54, CNM67, SPO21/MPC70, ADY3  
UN and ADY4. Probable component of a spindle pole body (SPB) complex  
UN composed of ADY3, SSPI, DON1, MPC54, SPO21/MPC70, NUD1 and CNM67.  
UN -!- SUBCELLULAR LOCATION: Localizes to the meiotic outer plaque of the  
UN SPB, at the end of the meiotic spindles.  
UN -!- PTM: Phosphorylated from S/G2 phase until the end of mitosis.  
UN -!- SIMILARITY: Contains 8 leucine-rich (LRR) repeats.  
UN -----  
UN This SWISS-PROT entry is copyright. It is produced through a collaboration  
UN between the European Bioinformatics Institute and the EMBL outstation -  
UN the European Bioinformatics Institute. There are no restrictions on its  
UN use by non-profit institutions as long as its content is in no way  
UN modified and this statement is not removed. Usage by and for commercial  
UN entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC EMBL; X62147; CAA44073.1; -  
DR EMBL; Z75281; CAA99704.1; -  
DR PIR; S67285; S67285  
DR GERMOnline; 143961; -  
DR SGD; S0005900; NUD1.  
DR GO; GO:0005816; C:spindle pole body; IDA.  
DR GO; GO:0005200; F:structural constituent of cytoskeleton; IPI.  
DR GO; GO:0007020; P:microtubule nucleation; IPI.  
DR InterPro; IPR001611; LRR.  
DR Pfam; PF00560; LRR; 3.  
KW Cell cycle; Mitosis; Meiosis; Repeat; Leucine-rich repeat.  
FT REPEAT 499 525 LRR 1.  
FT REPEAT 542 564 LRR 2.  
FT REPEAT 565 585 LRR 3.  
FT REPEAT 586 612 LRR 4.  
FT REPEAT 619 640 LRR 5.  
FT REPEAT 641 665 LRR 6.  
FT REPEAT 717 741 LRR 7.  
FT REPEAT 766 792 LRR 8.  
FT DOMAIN 223 248 ASN-RICH.  
FT DOMAIN 223 239 POLY-ASN.  
FT DOMAIN 266 274 POLY-SER.  
FT CONFLICT 5 6 TQ -> SE (IN REF. 2).  
SQ SEQUENCE 851 AA; 94103 MW; CB9F0408633C1315 CRC64;  
Query Match 44.4%; Score 44; DB 1; Length 851;  
Best Local Similarity 55.6%; Pred. No. 49;  
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
QY 2 KKGKGIIVKQNGKLEQ 19  
Db 674 KNTGALSCLKQNYKLDQ 691  
RESULT 14  
FRDA PROVV  
ID FRDA PROVV STANDARD; PRT; 598 AA.  
AC P20922;  
DT 01-PEB-1991 (Rel. 17, Created)  
DT 01-PEB-1991 (Rel. 17, Last sequence update)  
DT 28-PEB-2003 (Rel. 41, Last annotation update)  
DE Fumarate reductase flavoprotein subunit (EC 1.3.99.1).  
GN FRDA.  
OS Proteus vulgaris.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Proteus.  
OX NCBI\_TaxID=585;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88004470; PubMed=3308458;  
RA Cole S.T.;  
RT "Nucleotide sequence and comparative analysis of the frd operon  
RT encoding the fumarate reductase of Proteus vulgaris. Extensive  
RT sequence divergence of the membrane anchors and absence of an  
RT frd-linked ampC cephalosporinase gene."; Eur. J. Biochem. 167:481-488(1987)  
CC -!- FUNCTION: TWO DISTINCT, MEMBRANE-BOUND, FAD-CONTAINING ENZYMES ARE  
CC RESPONSIBLE FOR THE CATALYSIS OF FUMARATE AND SUCCINATE  
CC INTERCONVERSION; THE FUMARATE REDUCTASE IS USED IN ANAEROBIC  
CC GROWTH, AND THE SUCCINATE DEHYDROGENASE IS USED IN AEROBIC GROWTH.  
CC -!- CATALYTIC ACTIVITY: Succinate + acceptor = fumarate + reduced  
CC acceptor.  
CC -!- COFACTOR: FAD; covalently linked.  
CC -!- SUBUNIT: FUMARATE DEHYDROGENASE FORMS PART OF AN ENZYME COMPLEX  
CC CONTAINING FOUR SUBUNITS: A FLAVOPROTEIN, AN IRON-SULFUR, AND  
CC TWO HYDROPHOBIC ANCHOR PROTEINS.  
CC -!- SIMILARITY: Belongs to the FAD-dependent oxidoreductase family 2.  
CC FRD/SDH subfamily.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

un on: March 7, 2004, 13:31:46 ; Search time 7.30769 Seconds  
(without alignments)  
250.098 Million cell updates/sec

title: US-09-643-458b-12

effect score: 99

sequence: 1 CKGGGILVKQNGKLESQ 19

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 283366 seqs, 96191526 residues

total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

PIR 78:\*

1: piri:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	70	70.7	952	1 HXAD5	hexon protein - hu
2	51	51.5	224	2 H71506	probable S3 riboso
3	51	51.5	224	2 A81665	ribosomal protein
4	48	48.5	223	2 F86570	S3 ribosomal prote
5	48	48.5	223	2 C72055	S3 ribosomal prote
6	48	48.5	223	2 F81611	ribosomal protein
7	47	47.5	219	2 S35928	6-phosphofructokin
8	46	46.5	664	1 JX0336	succinate dehydrog
9	46	46.5	664	2 A84742	probable polygalac
10	46	46.5	665	1 A42792	succinate dehydrog
11	46	46.5	1272	2 T49313	copia-type reverse
12	46	46.5	1320	2 F96614	probable copia-typ
13	46	46.5	1352	2 F86246	hypothetical prote
14	46	46.5	1352	2 T47925	copia-type polypro
15	45	45.5	409	2 S63229	probable polyketid
16	45	45.5	603	1 VCFVER	env polypeptin -
17	44.5	44.9	596	2 AB1047	succinate dehydrog
18	44.5	44.9	602	1 RDECFE	fumarate reductase
19	44.5	44.9	602	2 G91270	flavoprotein subun
20	44.5	44.9	602	2 G86111	flavoprotein subun
21	44	44.4	71	2 F72222	ribosomal protein
22	44	44.4	114	2 C35720	hypothetical 12.4K
23	44	44.4	175	2 F64495	hypothetical prote
24	44	44.4	296	1 G71196	probable ribokinase
25	44	44.4	851	2 S67285	NUDI protein - yea
26	43	43.4	130	2 S77271	hypothetical prote
27	43	43.4	210	2 T06656	hypothetical prote
28	43	43.4	281	2 P69400	2-deoxy-D-gluconat
29	43	43.4	598	1 RDEBFV	fumarate reductase

30 43 43.4 599 2 H64097 fumarate reductase  
31 43 43.4 607 2 AH0044 succinate dehydrog  
32 43 43.4 1125 1 F70177 transcription-repa  
33 42.5 42.9 714 2 C81018 iron-regulated out  
34 42 42.4 258 2 A64564 flagellar export p  
35 42 42.4 258 2 G71946 probable flagellar  
36 42 42.4 528 2 S38242 hypothetical prote  
37 42 42.4 1418 2 S64918 hypothetical prote  
38 41.5 41.9 266 2 I40411 PDSX prophage ORF  
39 41 41.4 134 2 S51835 profilin - common  
40 41 41.4 187 2 T40047 hypothetical prote  
41 41 41.4 323 2 AC0308 probable O-succiny  
42 41 41.4 369 2 B72535 hypothetical prote  
43 41 41.4 416 2 G81258 serine transporter  
44 41 41.4 536 1 D70184 probable ribose/ga  
45 41 41.4 594 2 A75379 peptide ABC transp

#### ALIGNMENTS

##### RESULT 1

HXAD5

hexon protein - human adenovirus 5

C:Species: Mastadenovirus\_h5 (human adenovirus 5)

A:Note: host Homo sapiens (man)

C:Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 22-Oct-1999

C:Accession: A03849

R:Kinloch, R.; Mackay, N.; Mautner, V.

J. Biol. Chem. 259, 6431-6436, 1984

A:Title: Adenovirus hexon. Sequence comparison of subgroup C serotypes 2 and 5.

A:Reference number: A03849; MUID:84212465; PMID:6202684

A:Accession: A03849

A:Molecule type: DNA

A:Residues: 1-952 <KIN>

A:Cross-references: GB:X02997; GB:J01966; GB:J01980; GB:K02368; GB:V00029; GB:V00030; NI

C:Genetics:

A:Map position: 51.6-59.7

C:Superfamily: adenovirus hexon protein

C:Keywords: hexon protein

Query Match 70.7%; Score 70; DB 1; Length 952;  
Best Local Similarity 88.2%; Pred. No. 0.0044;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GKGGILVKQNGKLESQ 19

Db 245 GKGGILVKQNGKLESQ 261

##### RESULT 2

H71506

probable S3 ribosomal protein - Chlamydia trachomatis (serotype D, strain UW3/Cx)

C:Species: Chlamydia trachomatis

C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 13-Aug-1999

C:Accession: H71506

R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,

Science 282, 754-759, 1998

A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trach

A:Reference number: A71570; MUID:9900809; PMID:9784136

A:Accession: H71506

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-224 <ARN>

A:Cross-references: GB:AE001323; GB:AE001273; MID:G3328931; PIDN:AA68123.1; PID:G332895;

A:Experimental source: serotype D, strain UW-3/Cx

C:Genetics:

A:Gene: rs3

C:Superfamily: Escherichia coli ribosomal protein S3

Query Match 51.5%; Score 51; DB 2; Length 224;

Best Local Similarity 47.1%; Pred. No. 1.2;

Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

```
QY 1 CKGKGGLVKKQNGKLE 17
      ||| ||| ||| |||
Db 49 CQGAAGFVVRMSGKIE 65

RESULT 3
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606; PMID:10192388
A:Accession: C72055
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-223 <ARN>
A:Cross-references: GB:AE001647; GB:AE001363; NID:94376920; PIDN:AA18780.1; PID:943769;
A:Experimental source: strain CWL029
C:Genetics: rs3
C:Superfamily: Escherichia coli ribosomal protein S3

Query Match 48.5%; Score 48; DB 2; Length 223;
Best Local Similarity 41.2%; Pred. No. 3.7;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 CKGKGGLVKKQNGKLE 17
      ||| ||| ||| |||
Db 49 CQGAAGFVVRMSGKIE 65

RESULT 6
F81611
ribosomal protein S3 CP0106 [imported] - Chlamydothila pneumoniae (strain AR39)
C:Species: Chlamydothila pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
C:Accession: F81611
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: F81611
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-223 <REA>
A:Cross-references: GB:AE002173; GB:AE002161; NID:97189033; PIDN:AAF37989.1; PID:9718904
A:Experimental source: strain AR39, HL cells
C:Genetics:
C:Superfamily: Escherichia coli ribosomal protein S3

Query Match 48.5%; Score 48; DB 2; Length 223;
Best Local Similarity 41.2%; Pred. No. 3.7;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 CKGKGGLVKKQNGKLE 17
      ||| ||| ||| |||
Db 49 CQGAAGFVVRMSGKIE 65

RESULT 7
S3528
6-phosphofructokinase (EC 2.7.1.11) - Lactobacillus delbrueckii subsp. bulgaricus
C:Species: Lactobacillus delbrueckii subsp. bulgaricus
C:Date: 22-Nov-1993 #sequence_revision 03-Aug-1995 #text_change 12-May-2003
C:Accession: A48663; S16106; S3528
R:Branny, P.; De La Torre, F.; Garel, J.R.
J. Bacteriol. 175, 5344-5349, 1993
A:Title: Cloning, sequencing, and expression in Escherichia coli of the gene coding for
A:Reference number: A48663; MUID:93374827; PMID:8366023
A:Accession: A48663
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-319 <BRA>
A:Cross-references: EMBL:X71403; NID:g1154864; PIDN:CAA50526.1; PID:g396136
A:Note: authors translated the codon CAC for residue 161 as Arg, and CGC for residue 162
A:Note: submitted to the EMBL Data Library, April 1993
R:le Bras, G.; Deville-Bonne, D.; Garel, J.R.
Eur. J. Biochem. 198, 683-687, 1991
A:Title: Purification and properties of the phosphofructokinase from Lactobacillus bulgæ
A:Reference number: S16106; MUID:91266952; PMID:1828763
A:Accession: S16106
```



```

;Status: preliminary
;Molecule type: protein
;Residues: 1-28,'X',30-37,'Y' <UEA>
;Superfamily: 6-phosphofructokinase, bacterial type; 6-phosphofructokinase 1 homology
;Keywords: ATP; phosphotransferase
;4-278/Domain: 6-phosphofructokinase 1 homology <6PFI>

Query Match 47.5%; Score 47; DB 2; Length 319;
Best Local Similarity 52.3%; Pred. No. 7;
Matches 9; Conservative 4; Mismatches 0; Gaps 0;

>y 2 KKGKGGILVKQNGKLES 18
:||||:|:||||:|
>b 277 EKGGLAVGTEGNGKVT 293
:||||:|:||||:|

RESULT 8
succinate dehydrogenase (ubiquinone) (EC 1.3.5.1) flavoprotein chain precursor, mitochondr
;Contains: fumarate reductase (EC 1.3.99.1) flavoprotein
;Species: Homo sapiens (man)
;Date: 28-Oct-1994 #sequence_revision 19-Jul-1996 #text_change 16-Jun-2000
;Accession: JX0336; I52450
;Hirawake, H.; Wang, H.; Kuramochi, T.; Kojima, S.; Kita, K.
;Biochem. 116, 221-227, 1994
;Title: Human complex II (succinate-ubiquinone oxidoreductase): cDNA cloning of the fla
;Reference number: JX0336; MUID:95096020; PMID:7798181
;Accession: JX0336
;Molecule type: mRNA
;Residues: 1-664 <HR>
;Cross-references: DDBJ:D30648; NID:9506337; PIDN:BAA0332.1; PID:9506338
;Experimental source: liver
;Morris, A.A.; Farnsworth, L.; Ackrell, B.A.; Turnbull, D.M.; Birch-Machin, M.A.
;Biochim. Biophys. Acta 1185, 125-128, 1994
;Title: The cDNA sequence of the flavoprotein subunit of human heart succinate dehydrog
;Reference number: I52450; MUID:94190953; PMID:8142412
;Accession: I52450
;Status: translated from GB/EMBL/DDBJ
;Molecule type: mRNA
;Residues: 1-628,'P',630-656,'I',658-664 <MOR>
;Cross-references: GB:I21936; NID:G347133; PIDN:AAA20683.1; PID:G347134
;Genetics:
;Gene: GDB:SDH2
;Cross-references: GDB:378037; OMIM:600857
;Map position: Sp15-Sp15
;Complex: heterotetramer of two flavoprotein chains and two iron-sulfur protein chains
;Description: catalyzes the reversible oxidation of succinate to fumarate by ubiquinone
;Pathway: tricarboxylic acid cycle
;Superfamily: fumarate reductase flavoprotein; 3-oxosteroid 1-dehydrogenase homology; f
;Keywords: FAD; flavoprotein; heterotetramer; oxidoreductase; phosphoprotein; tri
;1-43/Domain: transit peptide (mitochondrion) #status predicted <TRP>
;44-664/Product: succinate dehydrogenase (ubiquinone) flavoprotein chain #status predic
;63-331/Domain: fumarate reductase flavoprotein homology <FRF>
;362-463/Domain: 3-oxosteroid 1-dehydrogenase homology <OXD>
;99/Modified site: 3'-FAD-histidine (His) #status predicted
;311/Active site: Cys #status predicted

Query Match 46.5%; Score 46; DB 1; Length 564;
Best Local Similarity 58.3%; Pred. No. 23;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

>y 1 CKKGKGGILVKQ 12
:||||:|:||||:|
>b 311 CRGEGGILINSQ 322
:||||:|:||||:|

RESULT 9
A84742
probable polygalacturonase [imported] - Arabidopsis thaliana
;Species: Arabidopsis thaliana (mouse-ear cress)
;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
;Accession: A84742

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; I
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanzaken, S.E.; Unayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84742
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-664 <STO>
A;Cross-references: GB:AE002093; NID:G2924778; PIDN:AAC04907.1; GSPDB:GN00139
A;Genetics:
A;Gene: At2g33160
A;Map position: 2

Query Match 46.5%; Score 46; DB 2; Length 664;
Best Local Similarity 53.3%; Pred. No. 23;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

>y 3 KKGKGGILVKQNGKLE 17
:||||:|:||||:|
>b 526 KKGKGGILVKQNGRYE 540
:||||:|:||||:|

RESULT 10
A42792
succinate dehydrogenase (ubiquinone) (EC 1.3.5.1) flavoprotein chain precursor, mitochondr
;Contains: fumarate reductase (EC 1.3.99.1) flavoprotein
;Species: Bos primigenius taurus (cattle)
;Date: 31-Dec-1993 #sequence_revision 19-Jul-1996 #text_change 09-Apr-1999
;Accession: A42792
;Birch-Machin, M.A.; Farnsworth, L.; Ackrell, B.A.; Cochran, B.; Jackson, S.; Bindoff, J
;J. Biol. Chem. 267, 11553-11558, 1992
A;Title: The sequence of the flavoprotein subunit of bovine heart succinate dehydrogenas
;Reference number: A42792; MUID:92283874; PMID:1375942
;Accession: A42792
A;Molecule type: mRNA; protein
A;Residues: 1-665 <BIR>
A;Cross-references: GB:M60879; NID:G387583
A;Experimental source: heart
A;Note: sequence extracted from NCBI backbone (NCBIP:104758)
A;Complex: heterotetramer of two flavoprotein chains and two iron-sulfur protein chains
A;Function:
A;Description: catalyzes the reversible oxidation of succinate to fumarate by ubiquinone
A;Pathway: tricarboxylic acid cycle
A;Superfamily: fumarate reductase flavoprotein; 3-oxosteroid 1-dehydrogenase homology; f
;Keywords: FAD; flavoprotein; heterotetramer; oxidoreductase; phosphoprotein; tri
;1-43/Domain: transit peptide (mitochondrion) #status predicted <TRP>
;44-664/Product: succinate dehydrogenase (ubiquinone) flavoprotein chain #status predic
;64-332/Domain: fumarate reductase flavoprotein homology <FRF>
;64-92/Region: beta-alpha-beta FAD nucleotide-binding fold
;F63-464/Domain: 3-oxosteroid 1-dehydrogenase homology <OXD>
;F100/Modified site: 3'-FAD-histidine (His) #status predicted
;F312/Active site: Cys #status predicted

Query Match 46.5%; Score 46; DB 1; Length 665;
Best Local Similarity 58.3%; Pred. No. 23;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

>y 1 CKKGKGGILVKQ 12
:||||:|:||||:|
>b 312 CRGEGGILINSQ 323
:||||:|:||||:|

RESULT 11
T49313
copia-type reverse transcriptase-like protein - Arabidopsis thaliana
N;Alternate names: protein T16L24.270
;Species: Arabidopsis thaliana (mouse-ear cress)
;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Sep-2000
;Accession: T49313
R;D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; L
submitted to the Protein Sequence Database, April 2000

```

A;Reference number: Z5017

A;Accession: T49313

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1272 <DNA>

A;Cross-references: ENBL:AL138659; GSPDB:GN00061; ATSP:T16L24.270

A;Experimental source: cultivar Columbia; BAC clone T16L24

C;Genetics:

A;Gene: ATSP:T16L24.270

A;Map position: 3

A;Introns: 1101/3

C;Superfamily: retrovirus-related polyprotein

Query Match 46.5%; Score 46; DB 2; Length 1272;

Best Local Similarity 61.5%; Pred. No. 44;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 KGKGGILVKQNG 14

|||||::: |||

Db 374 KGKGNILRLKNG 386

#### RESULT 12

F96614

probable copia-type polyprotein T1824.5 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cross)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001

C;Accession: F96614

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: F96614

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1320 <STO>

A;Cross-references: GB:AE005173; NID:g11038503; PIDN:AG27780.1; GSPDB:GN00141

C;Genetics:

A;Gene: T1824.5

A;Map position: 1

C;Superfamily: retrovirus-related polyprotein

Query Match 46.5%; Score 46; DB 2; Length 1320;

Best Local Similarity 61.5%; Pred. No. 46;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 KGKGGILVKQNG 14

|||||::: |||

Db 374 KGKGNILRLKNG 386

#### RESULT 13

F86246

hypothetical protein [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cross)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C;Accession: F86246

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: F86246

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1352 <STO>

A;Cross-references: GB:AE005172; NID:g5734736; PIDN:AA50001.1; GSPDB:GN00141

C;Genetics:

A;Map position: 1

C;Superfamily: retrovirus-related polyprotein

Query Match 46.5%; Score 46; DB 2; Length 1352;

Best Local Similarity 61.5%; Pred. No. 47;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 KGKGGILVKQNG 14

|||||::: |||

Db 374 KGKGNILRLKNG 386

#### RESULT 14

T47925

copia-type polyprotein - Arabidopsis thaliana

N;Alternate names: protein T20K12.230

C;Species: Arabidopsis thaliana (mouse-ear cross)

C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 19-May-2000

C;Accession: T47925

R;De Haan, M.; Maarse, A.C.; Grivell, L.A.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quei

submitted to the Protein Sequence Database, January 2000

A;Reference number: Z24480

A;Accession: T47925

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1352 <DEH>

A;Cross-references: EMBL:AL137898

A;Experimental source: cultivar Columbia; BAC clone T20K12

C;Genetics:

A;Map position: 3

A;Note: T20K12.230

C;Superfamily: retrovirus-related polyprotein

Query Match 46.5%; Score 46; DB 2; Length 1352;

Best Local Similarity 61.5%; Pred. No. 47;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 KGKGGILVKQNG 14

|||||::: |||

Db 374 KGKGNILRLKNG 386

#### RESULT 15

S69229

probable polyketide synthase 2 Snoa2 [similarity] - Streptomyces nogalater

C;Species: Streptomyces nogalater

C;Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 04-Mar-2000

C;Accession: S69229; S52401; T46677

R;Ylthonko, K.; Tuikka, J.; Jussila, S.; Cong, L.; Maentse, P.

Mol. Gen. Genet. 251, 113-120, 1996

A;Title: A gene cluster involved in nogalamycin biosynthesis from Streptomyces nogalater

A;Reference number: S69223; MUID:96242142; PMID:8668120

A;Accession: S69229

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-409 <YLI>

A;Cross-references: EMBL:Z48262; EMBL:AJ224512; PIDN:CAA12018.1

A;Experimental source: ATCC 27451

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995

C;Genetics:

A;Gene: sno2; snoa2

A;Start codon: GTG

C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I; 3-oxoacyl-[acyl-carrier-prot

F;23-401/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>

Query Match 45.5%; Score 45; DB 2; Length 409;  
Best Local Similarity 47.1%; Pred. No. 21;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Y 2 KGKGGILVKQNGKLES 18  
b 153 RGPFGVLVTEQAGGLDA 169

Search completed: March 7, 2004, 13:38:25  
Job time : 8.30769 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: March 7, 2004, 13:27:25 ; Search time 30.0659 seconds  
(without alignments)  
178.554 Million cell updates/sec

itle: US-09-643-458B-12

fect score: 99

quence: 1 CKGGKGLVKKQNGKLSQ 19

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

imum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : A\_Geneseq\_29Jan04:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003as:\*

7: Geneseq2003bs:\*

8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	70	70.7	189	2	Aaw79541 Adenoviru
2	70	70.7	338	6	Aac23323 Human Hu5
3	70	70.7	338	6	Abp56675 Human ade
4	70	70.7	952	2	Aaw79539 Adenoviru
5	68	68.7	14	2	Aaw79549 Adenoviru
6	63	63.6	952	2	Aaw63118 Human ade
7	51	51.5	219	2	Aav37699 Chlamydia
8	51	51.5	224	6	Abu27378 Protein e
9	48	48.5	223	2	Aav35261 Chlamydia
10	48	48.5	223	6	Abu26797 Protein e
11	47	47.5	484	6	Abj26348 Aspergill
12	46	46.5	194	5	Abb90935 Herpicida
13	46	46.5	532	6	Abm04803 Murine su
14	46	46.5	593	6	Aag79678 Human ENZ
15	46	46.5	616	4	Aau28049 Novel hum
16	46	46.5	664	5	Abb91985 Herpicida
17	46	46.5	664	6	Abv56952 Human SDH
18	46	46.5	723	4	Aau28237 Novel hum
19	45	45.5	409	3	Aay05776 Streptom
20	45	45.5	409	3	Aay77274 Streptom
21	45	45.5	409	3	Aay78826 Amino aci
22	44.5	44.9	297	4	Aag92712 C Glutami
23	44	44.4	231	5	Abg77343 Selected
24	44	44.4	251	5	Abj11245 Yeast sel
25	44	44.4	307	6	Abu18250 Protein e

## ALIGNMENTS

### RESULT 1

AAW79541

ID AAW79541 standard; protein; 189 AA.

XX AC AAW79541;

XX AC AAW79541;

DT 17-OCT-2003 (revised)

DT 11-JAN-1999 (first entry)

XX XX

DE Adenovirus serotype 5 loop 1 region.

XX XX

XX Adenovirus serotype 5; Ad5; hexon; coat protein; gene therapy; vector.

XX OS Human adenovirus type 5.

XX WO9840509-A1.

XX PD 17-SEP-1998.

XX PF 13-MAR-1998; 98WO-US005033.

XX PR 13-MAR-1997; 97US-00816346.

XX (CORR ) CORNELL RES FOUND INC.

XX (GENV-) GENVEC INC.

XX Crystal RG, Falck-Pedersen E, Gall J, Kovesdi I, Wickham TJ;

XX WPI; 1998-506738/43.

XX N-PSDB; AAV61503.

XX Chimeric adenovirus coat protein - useful in, e.g. vector for gene

XX transfer to treat inherited genetic diseases.

XX Claim 7; Page 75-76; 112pp; English.

XX This is the amino acid sequence of the loop 1 (11) region of the hexon

XX protein of the adenovirus serotype 5 (Ad5) coat protein (see also

XX AAW79539). It is encoded by a claimed DNA sequence (see AAV61503). The

XX invention provides a chimeric adenoviral coat protein, particularly a

XX chimeric adenovirus hexon protein, that has a decreased ability or

XX inability to be recognised by a neutralising antibody directed against

XX the corresponding wild-type adenovirus coat protein. The chimeric

XX adenoviral coat protein has a non-native amino acid sequence, especially

XX comprising a deletion of an internal hexon protein sequence, preferably a

XX hypervariable region or entire loop. DNA sequences (see AAV61502-23)

XX encoding claimed Ad2 and Ad5 chimeric coat proteins (see AAW79540-61) are

XX claimed. Also claimed are an adenovirus vector that comprises the

Abv53159 Protein s  
Abv58443 Drosophil  
Aag19315 Arabidops  
Aag40219 Arabidops  
Aag40218 Arabidops  
Aag19314 Arabidops  
Aag40217 Arabidops  
Aag19313 Arabidops  
Aag18137 S. epider  
Abv38493 Staphyloc  
Abm68790 Phototrab  
Abu43694 Protein e  
Abu19289 Protein e  
Abu43125 Protein e  
Abb84282 S. phaeoc  
Aau73008 Neisseria  
Abu51103 Helicobac  
Aae31517 Sweet che  
Abu51915 Helicobac  
Abu52230 Helicobac

CC chimeric adenovirus coat protein, a method of genetically modifying a  
 CC cell by contacting it with the vector, and a host cell that comprises the  
 CC chimeric adenovirus coat protein. The vector can be used for gene  
 CC transfer, for the treatment of inherited diseases. It can also be used to  
 CC render certain cells susceptible to the killing action of certain drugs,  
 CC or to study the effects of expression of specific genes in a given cell,  
 CC or tissue in vitro or in vivo. (Updated on 17-OCT-2003 to standardise OS  
 CC field)

XX SQ Sequence 189 AA;

Query Match 70.7%; Score 70; DB 2; Length 189;  
 Best Local Similarity 88.2%; Pred. No. 0.0098;  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GKGILVKQNGKLESQ 19  
 | | | | | | | | | | | | | | | | | | | |  
 DB 114 GGGILVKQNGKLESQ 130

RESULT 2  
 AAO23323  
 ID AAO23323 standard; protein; 338 AA.

XX AC AAO23323;

XX DT 17-OCT-2003 (first entry)

XX DE Human Hu5 protein, homologous to chimpanzee adenovirus hexon regions.

XX KW Simian adenovirus; genetic engineering; immunogenic; yellow fever; hexon;  
 KW cancer; epidemic gastroenteritis; Japanese encephalitis; vaccine; Hu5;  
 KW canine distemper; chlamydial infection; schistosomiasis; antibacterial;  
 KW antiparasitic; cytostatic; fungicidal; virucidal; human.

XX OS Homo sapiens.

XX PN WO2003046124-A2.

XX PD 05-JUN-2003.

XX PF 20-NOV-2002; 2002WO-US033645.

XX PR 21-NOV-2001; 2001US-0331951P.

XX PR 22-MAR-2002; 2002US-0366798P.

XX PA (UYPE-) UNIV PENNSYLVANIA.

XX PI Wilson JM, Gao G, Roy S;

XX DR WPI; 2003-505189/47.

XX PT New simian adenovirus nucleic acid sequence, useful for preparing a  
 PT composition for immunizing an animal against bacteria, virus, parasites  
 PT or cancer cell.

XX PS Disclosure; Fig 1; 306pp; English.

XX CC This invention relates to novel isolated simian adenovirus nucleic acid  
 CC sequences comprising Pan5, Pan6, Pan7, SV1, SV25 and SV39. Adenovirus is  
 CC a double stranded DNA virus used widely in genetic engineering  
 CC applications due to its high efficiency of gene transfer and large  
 CC transgene capacity by the deletion of the E1 region to allow for the  
 CC insertion of the gene of interest. Specifically, the present invention  
 CC describes the use of three major proteins that encode the icosahedral  
 CC capsid, namely hexon, penton and knobbed fibre from a simian adenovirus.  
 CC The vector constructs derived from simian rather than human do not elicit  
 CC an immune response, such that they are useful for vaccines, and  
 CC furthermore for repeat administration and titre boosting by a second  
 CC vaccination if required. Accordingly, transgenes inserted into the  
 CC vectors of the invention can be immunogenic and used to treat for  
 CC example, yellow fever, epidemic gastroenteritis, Japanese encephalitis,  
 CC canine distemper, chlamydial infections and schistosomiasis. They could

CC also treat genetic deficiencies and cancer, such that the simian  
 CC adenovirus nucleic acids of the invention along with a heterologous  
 CC transgene can be described variously as antibacterial, antiparasitic,  
 CC cytostatic, fungicidal or virucidal. This polypeptide sequence is the  
 CC hexon region of the human Hu5 protein, used for homology purposes in the  
 CC invention. NOTE: The present sequence (in figure 1) is not further  
 CC described and differs from that identified as SeqID 13 in the sequence  
 CC listing of the specification

XX SQ Sequence 338 AA;

Query Match 70.7%; Score 70; DB 6; Length 338;  
 Best Local Similarity 88.2%; Pred. No. 0.018;  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GKGILVKQNGKLESQ 19  
 | | | | | | | | | | | | | | | | | | | |  
 DB 121 GGGILVKQNGKLESQ 137

RESULT 3

ABP56675

ID ABP56675 standard; protein; 338 AA.

XX AC ABP56675;

XX DT 25-MAR-2003 (first entry)

XX DE Human adenovirus serotype 5 capsid hexon protein SEQ ID NO:41.

XX KW Chimpanzee C68 adenovirus; rapid screening; bacterial transformant;  
 KW simian adenovirus; cytostatic; antiparasitic; antimicrobial; vaccine;  
 KW adenoviral; capsid protein; hexon; penton; fibre protein; cancer;  
 KW hyperproliferative condition; psoriasis; infection.

XX OS Human adenovirus type 5.

XX PN WO2003000851-A2.

XX PD 03-JAN-2003.

XX PF 20-JUN-2002; 2002WO-US019735.

XX PR 22-JUN-2001; 2001US-0300501P.

XX PR 04-JUN-2002; 2002US-0385632P.

XX PA (UYPE-) UNIV PENNSYLVANIA.

XX PI Gao G, Wilson JM;

XX DR WPI; 2003-184043/18.

XX PT New C68 chimpanzee adenoviral capsid protein, useful for preparing a  
 PT composition for treating hyperproliferative conditions e.g., cancer or  
 PT psoriasis and as a vaccine against bacterial, fungal, viral or parasitic  
 PT infection.

XX PS Disclosure; Fig 4; 124pp; English.

XX CC The present invention describes a chimpanzee C68 adenoviral capsid  
 CC protein, which is substantially free of other viral proteins with which  
 CC it is naturally associated, comprising: (a) a hexon protein comprising  
 CC 513-amino acid sequence (see ABP56652); (b) a penton protein comprising  
 CC 534-amino acid sequence (see ABP56647); (c) a fibre protein comprising  
 CC 425-amino acid sequence (see ABP56662); or (d) a unique fragment of any  
 CC of (A)-(C) comprising 8-amino acid residues in length. Also described:  
 CC (1) a novel adenovirus serotype comprising a unique fragment of the C68  
 CC hexon protein fused to a heterologous adenovirus hexon peptide; (2) a  
 CC recombinant or pseudotyped adenovirus comprising a capsid of the novel  
 CC adenovirus serotype encapsidating a molecule for delivery to a target  
 CC cell; (3) an adenoviral capsid; (4) a pharmaceutical composition; (5)  
 CC rapid screening of recombinant constructs; and (6) a host cell. The  
 CC chimpanzee C68 adenoviral capsid protein has cytostatic, antiparasitic

C and antimicrobial activities, and can be used in vaccines. The chimpanzee  
C C68 adenoviral capsid protein is useful for preparing a composition for  
C treating hyperproliferative conditions e.g., cancer or peoriasis and as a  
C vaccine against bacterial, fungal, viral or parasitic infection. The  
C present sequence represents a human adenovirus capsid hexon protein which  
C is given in comparison with the chimpanzee C68 adenovirus hexon protein  
C in the exemplification of the present invention

X C Sequence 338 AA;

Query Match 70.7%; Score 70; DB 6; Length 338;  
Best Local Similarity 88.2%; Pred. No. 0.018;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Y 3 GKGGLVVKQNGKLESQ 19  
b 121 GQGGLVVKQNGKLESQ 137

RESULT 4  
AAW79539  
D AAW79539 standard; protein; 952 AA.

C AAW79539;

X 17-OCT-2003 (revised)

T 11-JAN-1999 (first entry)

E Adenovirus serotype 5 hexon protein.

W Adenovirus serotype 5; Ad5; hexon; coat protein; gene therapy; vector.

X Human adenovirus type 5.

H Key Location/Qualifiers

T 132..320

T /label= l1 loop

T /note= "Claim 7"

T 138..192

T /label= HVR1 region

T /note= "Claim 7"

T 188..194

T /label= HVR2 region

T /note= "Claim 7"

T 212..219

T /label= HVR3 region

T /note= "Claim 7"

T 248..261

T /label= HVR4 region

T /note= "Claim 7"

T 263..283

T /label= HVR5 region

T /note= "Claim 7"

T 305..316

T /label= HVR6 region

T /note= "Claim 7"

T 412..462

T /label= l2 loop

T /note= "Claim 7"

T 422..450

T /label= HVR7 region

T /note= "Claim 7"

T 951

T /label= Gln, His, Thr

T 952

T /label= Gln, His, Thr

X WO9840509-A1.

CX 17-SEP-1998.

CX 13-MAR-1998; 98WO-US005033.

CX

PR 13-MAR-1997; 97US-00816346.

XX (CORR ) CORNELL RES FOUND INC.

PA (GENV-) GENVEC INC.

XX Crystal RG, Falck-Pedersen E, Gall J, Kovessdi I, Wickham TJ;

XX WPI; 1998-506738/43.

XX N-PSDB; AAV61501.

XX Chimeric adenovirus coat protein - useful in, e.g. vector for gene

XX transfer to treat inherited genetic diseases.

XX Disclosure; Page 66-70; 112pp; English.

XX This is the amino acid sequence of the wild-type hexon protein of

XX adenovirus serotype 5 (Ad5). The invention provides a chimeric adenoviral

XX coat protein, particularly a chimeric adenovirus hexon protein, that has

XX a decreased ability or inability to be recognised by a neutralising

XX antibody directed against the corresponding wild-type adenovirus coat

XX protein. The chimeric adenoviral coat protein has a non-native amino acid

XX sequence, especially comprising a deletion of one or more regions of the

XX hexon protein, particularly a hypervariable region (HVR) of loop 1 or 2

XX of the hexon protein. DNA sequences (see AAV61502-23) encoding claimed

XX Ad2 and Ad5 chimeric coat proteins (see AAW79540-61) are claimed. Also

XX claimed are an adenovirus vector that comprises the chimeric adenovirus

XX coat protein, a method of genetically modifying a cell by contacting with

XX the vector, and a host cell that comprises the chimeric adenovirus coat

XX protein. The vector can be used for gene transfer, for the treatment of

XX inherited diseases (e.g. by carrying the cystic fibrosis (CF)

XX transmembrane conductance regulator cDNA to treat CF). It can also be

XX used to render certain cells susceptible to the killing action of certain

XX drugs, or to study the effects of expression of specific genes in a given

XX cell or tissue in vitro or in vivo. (Updated on 17-OCT-2003 to

XX standardise OS field)

XX SQ Sequence 952 AA;

Query Match 70.7%; Score 70; DB 2; Length 952;

Best Local Similarity 88.2%; Pred. No. 0.055;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GKGGLVVKQNGKLESQ 19

Db 245 GQGGLVVKQNGKLESQ 261

RESULT 5

AAW79549

ID AAW79549 standard; peptide; 14 AA.

XX AAW79549;

XX 17-OCT-2003 (revised)

DT 11-JAN-1999 (first entry)

XX Adenovirus serotype 5 hypervariable region HVR4.

XX Adenovirus serotype 5; Ad5; hexon; coat protein; gene therapy; vector.

XX Human adenovirus type 5.

OS WO9840509-A1.

XX 17-SEP-1998.

XX 13-MAR-1998; 98WO-US005033.

XX 13-MAR-1997; 97US-00816346.

XX (CORR ) CORNELL RES FOUND INC.

XX (GENV-) GENVEC INC.

XX

PI Crystal RG, Falck-Pedersen E, Gall J, Kovsed I, Wickham TJ;  
 XX WPI; 1998-506738/43.  
 DR N-PSDB; AAV61511.  
 XX Chimeric adenovirus coat protein - useful in, e.g. vector for gene  
 PT transfer to treat inherited genetic diseases.  
 XX Claim 7; Page 80; 112pp; English.  
 XX This is the amino acid sequence of hypervariable region HVR4 of the hexon  
 CC protein (see AAW79539) of adenovirus serotype 5 (Ad5). It is encoded by a  
 CC claimed DNA sequence (see AAV61511). The invention provides a chimeric  
 CC adenoviral coat protein, particularly a chimeric adenovirus hexon  
 CC protein, that has a decreased ability or inability to be recognised by a  
 CC neutralising antibody directed against the corresponding wild-type  
 CC adenovirus coat protein. The chimeric adenoviral coat protein has a non-  
 CC native amino acid sequence, especially comprising a deletion of an  
 CC internal hexon protein sequence, preferably a hypervariable region or  
 CC entire loop. DNA sequences (see AAV61502-23) encoding claimed Ad2 and Ad5  
 CC chimeric coat proteins (see AAW79540-61) are claimed. Also claimed are a  
 CC adenovirus vector that comprises the chimeric adenovirus coat protein, a  
 CC method of genetically modifying a cell by contacting it with the vector,  
 CC and a host cell that comprises the chimeric adenovirus coat protein. The  
 CC vector can be used for gene transfer, for the treatment of inherited  
 CC diseases. It can also be used to render certain cells susceptible to the  
 CC killing action of certain drugs, or to study the effects of expression of  
 CC specific genes in a given cell or tissue in vitro or in vivo. (Updated on  
 CC 17-OCT-2003 to standardise OS field)  
 XX SQ Sequence 14 AA;

Query Match 68.7%; Score 68; DB 2; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 0.0012;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GILVQONGKLESQ 19  
 DB 1 GILVQONGKLESQ 14

RESULT 6  
 AAW63118  
 ID AAW63118 standard; protein; 952 AA.  
 XX AC AAW63118;  
 XX DT 17-OCT-2003 (revised)  
 XX DT 24-NOV-1998 (first entry)  
 XX DE Human adenovirus 5 hexon.  
 XX KW Adenovirus 5; Ad5; hexon; vector; gene therapy; vaccine.  
 XX OS Human adenovirus type 5.  
 XX FH Key Location/Qualifiers  
 FT Misc-difference 88 /note= "encoded by GGG"  
 FT Region 142..322 /label= L1  
 FT FT /note= "loop 1"  
 FT Region 405..517 /label= L2  
 FT FT /note= "loop 2"  
 FT Region 502..558 /label= L4  
 FT FT /note= "loop 4"  
 FT Misc-difference 602 /note= "encoded by ACG"  
 XX XX WO9832842-A1.  
 XX XX

PD 30-JUL-1998.  
 XX 22-JAN-1998; 98WO-US001113.  
 XX 24-JAN-1997; 97US-00788674.  
 XX (GENE-) GENETIC THERAPY INC.  
 XX Roy S;  
 XX WPI; 1998-427938/36.  
 DR N-PSDB; AAV42656.  
 DR Modified adenoviral vaccine - comprises adenovirus with a hexon loop  
 PT region of a second viral serotype.  
 XX Disclosure; Fig 1Aa-e; 67pp; English.  
 XX This is the amino acid sequence of the adenovirus 5 hexon. A modified  
 CC adenovirus, in which the original adenovirus is of a first serotype  
 CC within a first subgenus, and in which, in the modified version, at least  
 CC a portion of at least one loop region of the hexon is removed and  
 CC replaced with at least a portion of at least one loop region of the hexon  
 CC of an adenovirus of a second serotype within a second subgenus, is new.  
 CC Also claimed is a method of providing a therapeutic effect in a host  
 CC comprising the administration of the adenovirus. The invention can be  
 CC used as a vaccine against adenoviruses or as a gene therapy vector. The  
 CC adenovirus 12 hexon amino acid sequence is also provided (see AAW63117).  
 CC (Updated on 17-OCT-2003 to standardise OS field)  
 XX SQ Sequence 952 AA;

Query Match 63.6%; Score 63; DB 2; Length 952;  
 Best Local Similarity 98.2%; Pred. No. 0.67;  
 Matches 15; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 5 GG--ILVQONGKLESQ 19  
 DB 245 GGQCILVQONGKLESQ 261

RESULT 7  
 AAY37699  
 ID AAY37699 standard; protein; 219 AA.  
 XX AC AAY37699;  
 XX DT 07-OCT-1999 (first entry)  
 XX DE Chlamydia trachomatis ribosomal polypeptide.  
 XX KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;  
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KW Bartholinitis; pneumopathy; venereal lymphogranulomatosis.  
 XX OS Chlamydia trachomatis.  
 XX FN WO9928475-A2.  
 XX PD 10-JUN-1999.  
 XX PF 27-NOV-1998; 98WO-IB001939.  
 XX PR 28-NOV-1997; 97FR-00015041.  
 XX PR 17-DEC-1997; 97FR-00016034.  
 XX PR 04-NOV-1998; 98US-0107077P.  
 XX PA (GEST ) GENSET.  
 XX PI Griffiths R;  
 XX WPI; 1999-371125/31.  
 XX XX

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are:

- (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense





S Disclosure; Page; 175pp; English.

X The invention relates to novel purified or isolated nucleic acids of

C essential genes of *Aspergillus fumigatus*. The isolated nucleic acids of

C the invention are used to treat or prevent infections by a pathogenic

C organism such as *A. fumigatus*, to treat a non-infectious disease in a

C subject (e.g. cancer), to prevent or contain contamination of an object

C by *A. fumigatus*, or to prevent or inhibit formation on a surface of a

C biofilm comprising *A. fumigatus*. The polynucleotides are useful for

C expressing recombinant protein for characterisation, screening or

C therapeutic use, as markers for host tissues in which the pathogenic

C organisms invade or reside, for comparing with the DNA sequence of *A.*

C *fumigatus* to identify duplicated genes or paralogues having the same or

C similar biochemical activity and/or function, for comparing with DNA

C sequences of other related or distant pathogenic organisms to identify

C potential orthologous essential or virulence genes, for selecting and

C making oligomers for attachment to a nucleic acid array for examination

C of expression patterns, for raising anti-protein antibodies, as an

C antigen to raise anti-DNA antibodies or to elicit another immune

C response, and for identifying polynucleotides encoding the other protein

C with which binding occurs or to identify inhibitors of the binding

C interaction. The polypeptides may be used to raise antibodies or to

C elicit immune response, as a reagent in assays designed to quantitatively

C determine levels of the protein in biological fluids, as a marker for

C host tissues in which pathogenic organism invade or reside, and to

C isolate correlative receptors or ligands in the case of virulence

C factors. This sequence represents a protein of one of the essential genes

C of *Aspergillus fumigatus* of the invention

X

X Sequence 484 AA;

Query Match 47.5%; Score 47; DB 6; Length 484;

Best Local Similarity 61.5%; Pred. No. 96;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

NY 1 CKKGKGVILVKQON 13

b 103 CEDGGGVILVKQVN 115

RESULT 12

BB90935

ID ABB90935 standard; protein; 194 AA.

AC ABB90935;

XX 31-MAY-2002 (first entry)

XX Herbicidally active polypeptide SEQ ID NO 146.

XX Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.

XX WO200210210-A2.

XX 07-FEB-2002.

XX 28-AUG-2001; 2001WO-EP009892.

XX 28-AUG-2001; 2001WO-EP009892.

XX (FARB ) BAYER AG.

XX Tietjen K, Weidner M;

XX WPI; 2002-269010/31.

XX Identifying plant target proteins for herbicidally active compounds,

XX comprising aligning and comparing nucleic acid or amino acid sequences

XX from plant with nucleic acid or amino acid sequences from non-plant

XX organisms.

PS Claim 5; SEQ ID NO 146; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins (ABB90790-ABB94016)

CC for herbicidally active compounds, comprising aligning and comparing

CC nucleic acid or amino acid sequences from plant with nucleic acid or

CC amino acid sequences from non-plant organisms using suitable search

CC parameters, where plant sequences having an E-value greater by a factor

CC of 3 than the E-value of most similar non-plant sequences are selected.

CC The polypeptides or nucleic acids encoding them are useful for

CC identifying modulators. The identified modulators are useful as

CC herbicides

XX

SQ Sequence 194 AA;

Query Match 46.5%; Score 46; DB 5; Length 194;

Best Local Similarity 61.5%; Pred. No. 51;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 KGKGGVILVKQNG 14

DB 57 KGKGNILILKNG 69

RESULT 13

ABM04803

ID ABM04803 standard; protein; 532 AA.

XX AC ABM04803;

XX 22-SEP-2003 (first entry)

XX Murine succinate dehydrogenase Pp subunit.

XX spinal cord; neuropathic pain; central sensitisation pain; pain;

XX analgesic; gene therapy.

XX Mus musculus.

XX EP1284298-A2.

XX 19-FEB-2003.

XX 26-JUL-2002; 2002EP-00255229.

XX 27-JUL-2001; 2001GB-00018354.

XX 07-FEB-2002; 2002GB-00002883.

XX (WARN ) WARNER LAMBERT CO.

XX Brooksbank RA, Dixon AK, Lee K, Pimock RD;

XX WPI; 2003-543489/52.

XX N-PSDB; ACF25345.

XX Use of an isolated gene sequence in the screening of compounds for

XX diagnosing or treating pain.

XX Claim 1; Page 83-84; 188pp; English.

XX The invention relates to a novel isolated gene sequence that is

XX downregulated in the spinal cord of a mammal in response to mechanically

XX distinct first and second models of neuropathic or central sensitisation

XX pain, useful in the screening of compounds for diagnosing or treating

XX pain. A protein encoded by a gene of the invention has analgesic

XX activity. A polynucleotide of the invention may have a use in gene

XX therapy. The gene sequence is useful for preparing a composition for

XX diagnosing or treating pain. The present sequence represents a protein

XX encoded by a gene of the invention

XX

SQ Sequence 532 AA;

Query Match 46.5%; Score 46; DB 6; Length 532;

Best Local Similarity 58.3%; Pred. No. 1.5e+02;

QY	1	CKGKGILVKKQ 12	Matches	7;	Conservative	3;	Mismatches	2;	Indels	0;	Gaps	0;	FT	/label= Fumarate reductase/succinate dehydrogenase
Db	238	CRGEGGILINSQ 249											FT	/note= "Identified by BLIMPS_BLOCKS"
RESULT 14													FT	/label= Flavoprotein subunit oxidoreductase
AAG79678													FT	/note= "Identified by BLAST_PRODOW"
ID		AAG79678 standard; protein; 583 AA.											FT	/label= "Potentially phosphorylated"
XX													FT	/note= "Potentially phosphorylated"
AC													FT	/label= FAD binding domain
XX													FT	/note= "Identified by HMW_PPFAM"
XX													FT	/note= "Potentially phosphorylated"
DT													FT	/note= "Potentially phosphorylated"
XX		19-FEB-2003 (first entry)											FT	/note= "Potentially phosphorylated"
DE		Human ENZM-8, incyte ID No: 5730123CD1.											FT	/label= Fumarate reductase/succinate dehydrogenase
XX													FT	/note= "Identified by BLIMPS_BLOCKS"
KW		ENZM; cardiovascular disorder; arteriovenous fistula; prostate;											FT	/note= "Potentially phosphorylated"
KW		atherosclerosis; hypertension; Raynaud's disease; aneurysm; cervix;											FT	/label= "Potentially phosphorylated"
KW		varicose vein; thrombophlebitis; congestive heart failure; brain; breast;											FT	/label= FAD-dependent pyridine nucleotide reductase
KW		angina pectoris; ischaemic; heart disease; autoimmune; inflammation;											FT	/note= "Identified by BLIMPS_PRINTS"
KW		acquired immunodeficiency syndrome; anaemia; asthma; Crohn's disease;											FT	/label= "Identified by BLIMPS_PRINTS"
KW		neurological disorder; epilepsy; Huntington's disease; dementia; stroke;											FT	/label= FAD-dependent pyridine nucleotide reductase
KW		Alzheimer's disease; Creutzfeldt-Jakob disease; multiple sclerosis;											FT	/note= "Identified by BLIMPS_PRINTS"
KW		cerebral palsy; Parkinson's disease; anxiety; infection; sarcoma;											FT	/label= "Identified by BLIMPS_PRINTS"
KW		metabolic disorder; Addison's disease; goitre; thymic dysplasia;											FT	/label= Pyridine nucleotide disulphide reductase
KW		pneumonia; hepatitis; influenza; immune deficiency; infertility;											FT	/note= "Identified by BLIMPS_PRINTS"
KW		severe combined immunodeficiency disease; reproduction; eye disorder;											FT	/label= "Identified by BLIMPS_PRINTS"
KW		endometriosis; prostatitis; Peyronie's disease; impotence; myeloma;											FT	/label= Fumarate reductase/succinate dehydrogenase
KW		glaucoma; ocular hypertension; cell proliferation; psoriasis; myeloma;											FT	/note= "Identified by BLAST_DOMO"
KW		polycythemia vera; cancer; adenocarcinoma; leukemia; lymphoma; melanoma;											FT	/note= "Potentially phosphorylated"
KW		succinate dehydrogenase flavoprotein subunit; enzyme.											FT	/label= Fumarate reductase/succinate dehydrogenase
OS		Homo sapiens.											FT	/note= "Identified by BLIMPS_BLOCKS"
XX													FT	/label= "Identified by BLAST_PRODOW"
XX													FT	/note= "Potentially phosphorylated"
XX													FT	/note= "Potentially phosphorylated"
XX													FT	/label= Fumarate reductase/succinate dehydrogenase
FT													FT	/note= "Identified by BLAST_PRODOW"
FT													FT	/note= "Potentially phosphorylated"
FT													FT	/note= "Potentially phosphorylated"
FT													FT	/label= Fumarate reductase/succinate dehydrogenase
FT													FT	/note= "Identified by BLAST_PRODOW"
FT													FT	/label= Pyridine nucleotide disulphide reductase
FT													FT	/note= "Identified by BLIMPS_PRINTS"
FT													FT	/label= Fumarate reductase/succinate dehydrogenase
FT													FT	/note= "Identified by BLIMPS_BLOCKS"
FT													FT	/label= Fumarate reductase/succinate dehydrogenase
FT													FT	/note= "Identified by BLIMPS_BLOCKS"
FT													FT	/note= "Potentially phosphorylated"
FT													FT	/label= Fumarate reductase/succinate dehydrogenase
FT													FT	/note= "Identified by BLIMPS_BLOCKS"
FT													FT	/note= "Potentially phosphorylated"
FT													FT	/label= Fumarate reductase/succinate dehydrogenase
FT													FT	/note= "Identified by HMW_PPFAM"
FT													FT	/note= "Potentially phosphorylated"
FT													FT	/note= "Potentially phosphorylated"
FT													FT	/label= Fumarate reductase/succinate dehydrogenase
FT													FT	/note= "Identified by BLAST_PRODOW"
FT													FT	/label= "Potentially phosphorylated"
FT													FT	/note= "Potentially phosphorylated"
FT													FT	/label= Fumarate reductase/succinate dehydrogenase
FT													FT	/note= "Identified by BLAST_PRODOW"
FT													FT	/label= "Potentially phosphorylated"
FT													FT	/note= "Potentially phosphorylated"



CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic  
CC reactions and conditions, such as asthma or other respiratory problems.  
CC In addition, (1) affects biorhythms or circadian cycles of rhythms,  
CC fertility, metabolism, catabolism, anabolism, storage or elimination of  
CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides  
CC analgesic effects or other pain reducing effects, immunoglobulin like  
CC activity and can act as an antigen in a vaccine composition to raise an  
CC immune response. AAU28020-AAU28395 represent novel human secreted protein  
XX amino acid sequences of the invention

SQ Sequence 616 AA;  
Query Match 46.5%; Score 46; DB 4; Length 616;  
Best Local Similarity 58.3%; Pred No. 1.8e+02;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 CKGKGGLVKQQ 12  
Db 263 CRGGGILINSQ 274

Search completed: March 7, 2004, 13:35:10  
Job time : 32.0659 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model  
un on: March 7, 2004, 13:31:46 ; Search time 13.4615 Seconds  
(without alignments)  
250.098 Million cell updates/sec

Title: US-09-643-458B-15  
Perfect score: 190  
Sequence: 1 CKGKGIVNTTLTKVKPTQENGWEKDATEFSDK 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78.8  
1: PIR1.8  
2: PIR2.8  
3: PIR3.8  
4: PIR4.8

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	165	86.8	952	1 HXADS	hexon protein - hu
2	59	31.1	493	2 S39299	hexon protein - hu
3	59	31.1	942	2 S39298	hexon protein - hu
4	56	29.5	167	2 G89802	hypothetical prote
5	56	29.5	323	2 T27640	hypothetical prote
6	56	29.5	1280	2 T00365	hypothetical prote
7	55	28.9	1764	2 T15171	hypothetical prote
8	55	28.9	2116	2 T49818	glutamate synthase
9	54	28.4	317	2 B37804	fecR protein - Esc
10	54	28.4	412	2 T02725	probable serine/th
11	53	27.9	454	2 T21358	hypothetical prote
12	53	27.9	525	2 T21357	hypothetical prote
13	53	27.9	677	2 S73798	MG260 homolog H91
14	52	27.4	502	2 H64917	probable membrane
15	52	27.4	502	2 H90918	hypothetical prote
16	52	27.4	502	2 B85767	hypothetical prote
17	51.5	27.1	447	2 S39296	hexon protein - hu
18	51.5	27.1	474	1 GJMS11	Ig gamma-2b chain
19	51.5	27.1	936	2 S57637	hexon protein - hu
20	51.5	27.1	1027	2 H90861	probable multidrug
21	51.5	27.1	1027	2 H90861	probable efflux pu
22	51.5	27.1	1267	1 MWXR31	lambda 3 protein -
23	51	26.8	560	2 T05278	hypothetical prote
24	51	26.8	879	2 AC2347	hypothetical prote
25	50.5	26.6	382	2 S45552	serine-type D-Ala-
26	50.5	26.6	464	2 B71123	hypothetical prote
27	50.5	26.6	751	1 MWXRGB	probable core prot
28	50.5	26.6	1078	2 D87647	hypothetical prote
29	50.5	26.6	1757	2 T05204	hypothetical prote

30	50	26.3	285	2	C72269	hypothetical prote
31	50	26.3	376	2	H69038	heat shock protein
32	50	26.3	410	1	S28615	serine/threonine/t
33	50	26.3	534	2	S44886	ZK112.1 protein -
34	50	26.3	865	2	S77440	hypothetical prote
35	50	26.3	1498	2	B86302	hypothetical prote
36	50	26.3	1532	2	A61262	collagen alpha 1(X
37	49.5	26.1	282	2	T31941	hypothetical prote
38	49.5	26.1	387	2	T23915	hypothetical prote
39	49	25.8	139	2	C83597	hypothetical prote
40	49	25.8	145	2	AH2320	hypothetical prote
41	49	25.8	658	2	S38890	dnak-type molecula
42	49	25.8	707	1	TSBRAB	tryptophan synthas
43	49	25.8	906	2	AB2762	DNA gyrase subunit
44	49	25.8	930	2	A97543	(y09073) DNA Gyras
45	48.5	25.5	309	2	S63132	hypothetical prote

ALIGNMENTS

RESULT 1

HXADS  
hexon protein - human adenovirus 5  
C:Species: Mastadenovirus h5 (human adenovirus 5)  
A:Note: host Homo sapiens (man)  
C>Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 22-Oct-1999  
C:Accession: A03849  
R:Kinloch, R.; Mackay, N.; Mautner, V.  
J. Biol. Chem. 259, 6431-6436, 1984  
A>Title: Adenovirus hexon. Sequence comparison of subgroup C serotypes 2 and 5.  
A:Reference number: A03849; MUID:84212465; PMID:6202684  
A:Accession: A03849  
A:Molecule type: DNA  
A:Residues: 1-952 <NIN>  
A:Cross-references: GB:X02997; GB:J01966; GB:J01980; GB:K02368; GB:V00029; GB:V00030; NI  
C:Genetics:  
A:Map position: 51.6-59.7  
C:Superfamily: adenovirus hexon protein  
C:Keywords: hexon protein

Query Match 86.8%; Score 165; DB 1; Length 952;  
Best Local Similarity 100.0%; Pred. No. 3.5e-14;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	5	GVINTETLTQVKPTQENGWEKDATEFSDK	35
DB	419	GVINTETLTQVKPTQENGWEKDATEFSDK	449

RESULT 2

S39299  
hexon protein - human adenovirus 3 (fragment)  
C:Species: Mastadenovirus h3 (human adenovirus 3)  
C>Date: 19-Mar-1997 #sequence\_revision 06-Jun-1997 #text\_change 26-Aug-1999  
C:Accession: S39299  
R:Pring-Akerblom, P.; Adrian, T.  
submitted to the EMBL Data Library, November 1993  
A:Reference number: S39296  
A:Accession: S39299  
A:Molecule type: DNA  
A:Residues: 1-493 <PRI>  
A:Cross-references: EMBL:X76552; NID:9434911; PIDN:CAA54054.1; PID:9434912  
C:Superfamily: adenovirus hexon protein

Query Match 31.1%; Score 59; DB 2; Length 493;  
Best Local Similarity 48.0%; Pred. No. 4.6;  
Matches 12; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY	5	GVINTETLTQVKPTQENGWEKDA	29
DB	312	GIGPOHTYGGIKVKTDDTNGWEKDA	336

```
RESULT 3
S39298
hexon protein - human adenovirus 3
C:Species: Mastadenovirus h3 (human adenovirus 3)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Sep-1997
C:Accession: S39298
R:Pring-Akerblom, P.; Adrian, T.
submitted to the EMBL Data Library, November 1993
A:Reference number: S39296
A:Accession: S39298
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-942 <PRI>
A:Cross-references: EMBL:X76549; NID:G4343909; PID:gl220140
C:Superfamily: adenovirus hexon protein

Query Match 31.1%; Score 59; DB 2; Length 942;
Best Local Similarity 48.0%; Pred. No. 9;
Matches 12; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 5 GVINTEITLTKVKPKTQENGWEKDA 29
Db 411 GIGPGHTTQGIKVKTDNTNGWEKDA 435

RESULT 4
G89802
hypothetical protein esb [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: G89802
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiranatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: G89802
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-167 <KUR>
A:Cross-references: GB:BA000018; PID:gl3700280; PIDN:BA841578.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: esb
C:Superfamily: bacterial single-stranded DNA-binding protein; single-stranded DNA-binding

Query Match 29.5%; Score 56; DB 2; Length 167;
Best Local Similarity 32.4%; Pred. No. 3.7;
Matches 12; Conservative 8; Mismatches 13; Indels 4; Gaps 1;

QY 2 KKGVINTEITLTK-----VKPKTQENGWEKDATEPSD 34
Db 86 EGRRFVTEVVCDSVQFLEPKNAQQGQGRQNEFOD 122

RESULT 5
T27640
hypothetical protein ZK1010.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T27640
R:Gardner, A.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z20398
A:Accession: T27640
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-323 <WIL>
A:Cross-references: EMBL:Z82083; PIDN:CA804969.1; GSPDB:GN00021; CESP:ZK1010.3
A:Experimental source: clone ZK1010
C:Genetics:
```

```
A:Gene: CESP:ZK1010.3
A:Map position: 3
A:Introns: 45/2; 72/2; 242/3
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK1010.3

Query Match 29.5%; Score 56; DB 2; Length 323;
Best Local Similarity 38.2%; Pred. No. 7.5;
Matches 13; Conservative 3; Mismatches 18; Indels 0; Gaps 0;

QY 2 KKGVINTEITLTKVKPKTQENGWEKDATEPSDK 35
Db 77 KBKKKKKKDDKEKIDPDITVGGGWRKIADDFDMK 110

RESULT 6
T00365
hypothetical protein KIAA0670 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00365
R:Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.
DNA Res. 5, 169-176, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete
A:Reference number: Z14142; MUID:98403880; PMID:9734811
A:Accession: T00365
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1280 <ISH>
A:Cross-references: EMBL:AB014570; NID:G3327153; PIDN:BAA31645.1; PID:G3327154
A:Experimental source: brain; clone HK02359
C:Genetics:
A:Note: KIAA0670

Query Match 29.5%; Score 56; DB 2; Length 1280;
Best Local Similarity 42.9%; Pred. No. 32;
Matches 12; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 4 KGVINTEITLTKVKPKTQENGWEKDATE 31
Db 852 KGLKICRTVTQVPAEGQENGQREBE 879

RESULT 7
T15171
hypothetical protein C48E7.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15171
R:Wansley, P.; Kramer, J.
submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid C48E7.
A:Reference number: Z18303
A:Accession: T15171
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1764 <WAM>
A:Cross-references: EMBL:AF000262; NID:gl947126; PID:gl947133; PIDN:AAB52938.1; GSPDB:GN
A:Experimental source: strain Bristol N2; clone C48E7
C:Genetics:
A:Gene: CESP:C48E7.6
A:Map position: 1
A:Introns: 27/3; 74/3; 190/2; 235/2; 305/3; 438/3; 464/3; 570/3; 645/2; 688/1; 765/3; 82

Query Match 28.9%; Score 55; DB 2; Length 1764;
Best Local Similarity 39.1%; Pred. No. 61;
Matches 18; Conservative 4; Mismatches 8; Indels 16; Gaps 3;

QY 6 VINTETL-----TKVKPKTQENG---W---EKDATEPSDK 35
Db 1399 VINSEHLGAFSGMGNRSTLKYKITSPENGTFYVWAGEKEAFESQK 1444

RESULT 8
```

A:Map position: 93 min  
C:Function:  
A>Description: is proposed to be the sensor that recognizes iron(III)citrate in the periplasmic space  
C:Superfamily: Fe2+-dicitrate sensor, transmembrane component  
C:Keywords: periplasmic space

Query Match 28.4%; Score 54; DB 2; Length 317;  
Best Local Similarity 47.6%; Pred. No. 14;  
Matches 10; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

OY 15 VKPKTGOENGWEKATERSDK 35  
||| : |||  
Db 232 VPLDDESTWTKILSFSDK 252  
||| : |||

RESULT 10  
T02725  
probable serine/threonine/tyrosine-specific protein kinase (EC 2.7.1.1-) T9I4.1 - Arabidopsis thaliana  
N:Alternate names: protein T9I4.1  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 16-Feb-2001  
C:Accession: T02725; F84690  
R:Rounslev, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.; Rounslev, S.D.; Cronin, L.A.; Sheen, M.; Vanaken, S.E.; Umayam, L.; Tallon, J.L.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.C.; Nierman, W.C.; White, O.; Eisele, D.J.  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: T02725  
A:Molecule type: DNA  
A>Status: translated from GB/EMBL/DDBJ  
A:Cross-references: EMBL:AC005315; NID:g3461834; PIDN:AAC33221.1; PID:g3461835; GSPDB:GN00139  
A:Experimental source: cultivar Columbia  
A:Residues: 1-412 <ROU>  
A:Residues: 1-412 <STO>  
A:Cross-references: GB:AE002093; NID:g3461835; PIDN:AAC33221.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: NTSP-T9I4.1; At2G28930  
A:Map position: 2  
A:Introns: 15/1; 121/3; 167/1; 214/3; 256/1  
C:Superfamily: Kinase-related transforming protein; protein kinase homology  
C:Keywords: ATP; phosphotransferase; serine/threonine/tyrosine-specific protein kinase  
F:67-358/Domain: protein kinase homology <KIN>

Query Match 28.4%; Score 54; DB 2; Length 412;  
Best Local Similarity 38.7%; Pred. No. 18;  
Matches 12; Conservative 5; Mismatches 6; Indels 8; Gaps 1;

OY 4 KGVINTEITLVKKPKTG -----QENGWE 26  
||| : |||  
Db 85 KGWIDEQTILTASKEGTGVVIAVKLNQDGWQ 115  
||| : |||

RESULT 11  
T21358  
hypothetical protein P25H5.1b - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T21358  
R:Steward, C.  
submitted to the EMBL Data Library, October 1996  
A:Reference number: Z19412  
A:Accession: T21358  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-454 <WIL>



A;Cross-references: EMBL:Z81068; PIDN:CAB02981.1; GSPDB:GNO0019; CESP:F25H5.1b  
A;Experimental source: clone F25H5

C;Genetics:

A;Gene: CESP:F25H5.1b

A;Map position: 1

A;Introns: 35/3; 114/3; 153/3; 186/3; 253/3; 288/1; 312/1

Query Match 27.9%; Score 53; DB 2; Length 454;

Best Local Similarity 33.3%; Pred. No. 27;

Matches 8; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 2 KGGVINTETLTKVKPKTQGENGW 25

Db 118 KCSGILETNEAVIAPKLGDSGTG 141

RESULT 12

T21357

hypothetical protein F25H5.1a - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T21357

R;Steward, C.

submitted to the EMBL Data Library, October 1996

A;Reference number: Z19412

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Accession: T21357

A;Molecule type: DNA

A;Residues: 1-525 <WIL>

A;Cross-references: EMBL:Z81068; PIDN:CAB02980.1; GSPDB:GNO0019; CESP:F25H5.1a

A;Experimental source: clone F25H5

C;Genetics:

A;Gene: CESP:F25H5.1a

A;Map position: 1

A;Introns: 35/3; 114/3; 153/3; 186/3; 253/3; 288/1; 312/1; 431/1; 478/3

Query Match

Best Local Similarity 27.9%; Score 53; DB 2; Length 525;

Matches 8; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 2 KGGVINTETLTKVKPKTQGENGW 25

Db 118 KCSGILETNEAVIAPKLGDSGTG 141

RESULT 13

S73798

M3260 homolog H91\_oxf677 - Mycoplasma pneumoniae (strain ATCC 29342)

C;Species: Mycoplasma pneumoniae

A;Variety: ATCC 29342

C;Date: 27-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 07-Dec-1999

C;Accession: S73798

R;Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.

Nucleic Acids Res. 24, 4420-4449, 1996

A;Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae

A;Reference number: S73327; MUID:97105885; PMID:8948633

A;Accession: S73798

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-677 <HIM>

A;Cross-references: EMBL:AB000047; GB:U00089; NID:g1674162; PIDN:AAB96120.1; PID:g167416

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996

C;Genetics:

A;Genetic code: SGC3

C;Superfamily: hypothetical protein MG185

Query Match

Best Local Similarity 27.9%; Score 53; DB 2; Length 577;

Matches 12; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

QY 2 KGGVINTETLTKVKPKTQGENGWKDATEFSDK 35

Db 116 KKGTEKTTTLTKTSSQSSNTTKDGKATSD 149

RESULT 14

H64917

probable membrane protein ydga - Escherichia coli (strain K-12)

C;Species: Escherichia coli

C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C;Accession: H64917

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: H64917

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-502 <BLAT>

A;Cross-references: GB:AE000257; GB:U00096; NID:g1787898; PIDN:AAC74686.1; PID:g1787900.

A;Experimental source: strain K-12, substrain MG1655

C;Genetics:

A;Gene: ydga

C;Superfamily: hypothetical protein b1614

C;Keywords: nucleotide binding; P-loop

F;5-21/Domain: transmembrane #status predicted <TM01>

F;256-263/Region: nucleotide-binding motif A (P-loop)

Query Match

Best Local Similarity 27.4%; Score 52; DB 2; Length 502;

Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 4 KGVINTETLTKVKPKTQGENGW 27

Db 62 RGVFSSQLQLLVKPIAGKPNFIK 85

RESULT 15

H90918

hypothetical protein ECs2320 [imported] - Escherichia coli (strain O157:H7, substrain R)

C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001

C;Accession: H90918

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A;Reference number: A99629; MUID:21156231; PMID:11258796

A;Accession: H90918

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-502 <HAY>

A;Cross-references: GB:BA000007; PIDN:BA035743.1; PID:g13361787; GSPDB:GNO0154

A;Experimental source: strain O157:H7, substrain RMD 050952

C;Genetics:

A;Gene: ECs2320

C;Superfamily: hypothetical protein b1614

Query Match

Best Local Similarity 27.4%; Score 52; DB 2; Length 502;

Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 4 KGVINTETLTKVKPKTQGENGW 27

Db 62 RGVFSSQLQLLVKPIAGKPNFIK 85

Search completed: March 7, 2004, 13:38:27

Job time : 15.4615 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

run on: March 7, 2004, 13:27:25 ; Search time 55.3846 Seconds  
(without alignments)  
178.554 Million cell updates/sec

title: US-09-643-458b-15

effect score: 190  
sequence: 1 CKGKGVINTLTKVKPTQNGWEKDATEPSDK 35

scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

minimum DB seq length: 0  
maximum DB seq length: 2000000000

post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

database : A\_Geneseq\_29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2002s:\*
- 6: Geneseq2003as:\*
- 7: Geneseq2003bs:\*
- 8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	165	86.8	51	2	Aaw79557 Adenoviru
2	165	86.8	338	6	Aao23323 Human Hu5
3	165	86.8	338	6	Abp56675 Human ade
4	165	86.8	952	2	Aaw79539 Adenoviru
5	165	86.8	952	2	Aaw63118 Human ade
6	151	79.5	29	2	Aaw79559 Adenoviru
7	56.5	29.7	93	5	Abb79316 Human ova
8	56	29.5	151	2	Aaw79358 DNA encod
9	56	29.5	167	4	Aau37429 Staphyloc
10	56	29.5	167	4	Aau34248 Staphyloc
11	56	29.5	167	4	Aau36819 Staphyloc
12	56	29.5	167	6	Abu16285 Protein e
13	56	29.5	167	6	Abm71695 Staphyloc
14	56	29.5	506	5	Abp41530 Human ova
15	56	29.5	519	4	Abg09940 Novel hum
16	56	29.5	567	2	Aaw8788 Polypepti
17	56	29.5	567	4	Abb56625 Human sec
18	56	29.5	567	6	Abu44882 Novel hum
19	56	29.5	567	7	Abu26362 Protein a
20	56	29.5	568	3	Aay85658 Human aci
21	56	29.5	583	3	Aay85659 Human aci
22	56	29.5	614	4	Abg09939 Novel hum
23	56	29.5	1280	6	Abu52985 Human put
24	56	29.5	1341	3	Aay85657 Human aci
25	56	29.5	2097	4	Abg09944 Novel hum

26	54	28.4	216	4	Abg30168 Novel hum
27	54	28.4	317	4	Aau34877 E. coli c
28	54	28.4	317	6	Abu15073 Protein e
29	54	28.4	323	6	Abp56670 Human ade
30	54	28.4	392	3	Aag29453 Arabidops
31	54	28.4	412	3	Aag29452 Arabidops
32	54	28.4	643	4	Abg30169 Novel hum
33	54	28.4	1610	4	Abb63784 Drosophil
34	53.5	28.2	921	6	Aao23317 Rhesus mo
35	53.5	28.2	924	4	Aag85051 Shrimp wh
36	53.5	28.2	1722	6	Abm67440 Phototrab
37	53.5	28.2	1748	6	Abm70311 Phototrab
38	53	27.9	702	5	Abp73923 Candida a
39	53	27.9	767	4	Abg11534 Novel hum
40	53	27.9	1342	4	Abg09943 Novel hum
41	52.5	27.6	89	4	Aab73746 Cysteine
42	52	27.4	422	2	Aar75389 Bacillus
43	52	27.4	663	4	Abg05177 Novel hum
44	52	27.4	1037	6	Abra41517 Human DIT
45	52	27.4	1765	6	Ada13339 Human int

ALIGNMENTS

RESULT 1  
AAW79557  
ID AAW79557 standard; protein; 51 AA.  
XX  
AC AAW79557;  
XX  
DT 17-OCT-2003 (revised)  
DT 11-JAN-1999 (first entry)  
XX  
DE Adenovirus serotype 5 loop 2 region.  
XX  
KW Adenovirus serotype 2; Ad2; hexon; coat protein; gene therapy; vector.  
XX  
OS Human adenovirus type 5.  
XX  
PN WO9840509-A1.  
XX  
PD 17-SEP-1998.  
XX  
PF 13-MAR-1998; 98WO-US005033.  
XX  
PR 13-MAR-1997; 97US-00816346.  
XX  
(CORR ) CORNELL RES FOUND INC.  
(GENV-) GENVEC INC.  
XX  
Crystal RG, Falck-Pedersen E, Gall J, Kovesdi I, Wickham TJ;  
WPI; 1998-506738/43.  
N-PSDB; AAV61519.  
XX  
Chimeric adenovirus coat protein - useful in, e.g. vector for gene transfer to treat inherited genetic diseases.  
XX  
Claim 7; Page 85; 112pp; English.  
XX  
This is the amino acid sequence of the loop 2 (12) region of the hexon protein of the adenovirus serotype 5 (Ad5) coat protein (see also AAW79539). It is encoded by a claimed DNA sequence (see AAV61519). The invention provides a chimeric adenoviral coat protein, particularly a chimeric adenovirus hexon protein, that has a decreased ability or inability to be recognised by a neutralising antibody directed against the corresponding wild-type adenovirus coat protein. The chimeric adenoviral coat protein has a non-native amino acid sequence, especially comprising a deletion of an internal hexon protein sequence, preferably a hypervariable region or entire loop. DNA sequences (see AAV61502-23) encoding claimed Ad2 and Ad5 chimeric coat proteins (see AAW79540-61) are claimed. Also claimed are an adenovirus vector that comprises the

CC chimeric adenovirus coat protein, a method of genetically modifying a  
 CC cell by contacting it with the vector, and a host cell that comprises the  
 CC chimeric adenovirus coat protein. The vector can be used for gene  
 CC transfer, for the treatment of inherited diseases. It can also be used to  
 CC render certain cells susceptible to the killing action of certain drugs,  
 CC or to study the effects of expression of specific genes in a given cell  
 CC or tissue in vitro or in vivo. (Updated on 17-OCT-2003 to standardise OS  
 CC field)

XX SQ Sequence 51 AA;

Query Match 86.8%; Score 165; DB 2; Length 51;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-16;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GVINTETTLTKVKPKTGQENGWEKDATEFSK 35  
 |||||  
 Db 8 GVINTETTLTKVKPKTGQENGWEKDATEFSK 38

RESULT 2

AAO23323  
 ID AAO23323 standard; protein; 338 AA.

XX AC AAO23323;

XX DT 17-OCT-2003 (first entry)

XX DE Human Hu5 protein, homologous to chimpanzee adenovirus hexon regions.

XX KW Simian adenovirus; genetic engineering; immunogenic; yellow fever; hexon;  
 KW cancer; epidemic gastroenteritis; Japanese encephalitis; vaccine; Hu5;  
 KW canine distemper; chlamydial infection; schistosomiasis; antibacterial;  
 KW antiparasitic; cytostatic; fungicidal; virucidal; human.

XX OS Homo sapiens.

XX PN WO2003046124-A2.

XX PD 05-JUN-2003.

XX PF 20-NOV-2002; 2002WO-US033645.

XX PR 21-NOV-2001; 2001US-0331951P.

XX FR 22-MAR-2002; 2002US-0366798P.

XX PA (TYPE-) UNIV PENNSYLVANIA.

XX PI Wilson JM, Gao G, Roy S;

XX DR WPI; 2003-505189/47.

XX PT New simian adenovirus nucleic acid sequence, useful for preparing a  
 PT composition for immunizing an animal against bacteria, virus, parasites  
 PT or cancer cell.

XX PS Disclosure; Fig 1; 306pp; English.

XX CC This invention relates to novel isolated simian adenovirus nucleic acid  
 CC sequences comprising Pan5, Pan6, Pan7, SV1, SV25 and SV35. Adenovirus is  
 CC a double stranded DNA virus used widely in genetic engineering  
 CC applications due to its high efficiency of gene transfer and large  
 CC transgene capacity by the deletion of the E1 region to allow for the  
 CC insertion of the gene of interest. Specifically, the present invention  
 CC describes the use of three major proteins that encode the icosahedral  
 CC capsid, namely hexon, penton and knobbed fibre from a simian adenovirus.  
 CC The vector constructs derived from simian rather than human do not elicit  
 CC an immune response, such that they are useful for vaccines, and  
 CC furthermore for repeat administration and titre boosting by a second  
 CC vaccination if required. Accordingly, transgenes inserted into the  
 CC vectors of the invention can be immunogenic and used to treat for  
 CC example, yellow fever, epidemic gastroenteritis, Japanese encephalitis,  
 CC canine distemper, chlamydial infections and schistosomiasis. They could

CC also treat genetic deficiencies and cancer, such that the simian  
 CC adenovirus nucleic acids of the invention along with a heterologous  
 CC transgene can be described variously as antibacterial, antiparasitic,  
 CC cytostatic, fungicidal or virucidal. This polypeptide sequence is the  
 CC hexon region of the human Hu5 protein, used for homology purposes in the  
 CC invention. NOTE: The present sequence (in figure 1) is not further  
 CC described and differs from that identified as SeqID 13 in the sequence  
 CC listing of the specification

XX SQ Sequence 338 AA;

Query Match 86.8%; Score 165; DB 6; Length 338;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-15;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GVINTETTLTKVKPKTGQENGWEKDATEFSK 35  
 |||||  
 Db 295 GVINTETTLTKVKPKTGQENGWEKDATEFSK 325

RESULT 3

ABP56675

ID ABP56675 standard; protein; 338 AA.

XX AC ABP56675;

XX DT 25-MAR-2003 (first entry)

XX DE Human adenovirus serotype 5 capsid hexon protein SEQ ID NO:41.

XX KW Chimpanzee C68 adenovirus; rapid screening; bacterial transformant;  
 KW simian adenovirus; cytostatic; antiparasitic; antimicrobial; vaccine;  
 KW adenoviral; capsid protein; hexon; penton; fibre protein; cancer;  
 KW hyperproliferative condition; psoriasis; infection.

XX OS Human adenovirus type 5.

XX PN WO2003000851-A2.

XX PD 03-JAN-2003.

XX PF 20-JUN-2002; 2002WO-US019735.

XX PR 22-JUN-2001; 2001US-0300501P.

XX FR 04-JUN-2002; 2002US-0385632P.

XX PA (TYPE-) UNIV PENNSYLVANIA.

XX PI Gao G, Wilson JM;

XX DR WPI; 2003-184043/18.

XX PT New C68 chimpanzee adenoviral capsid protein, useful for preparing a  
 PT composition for treating hyperproliferative conditions e.g., cancer or  
 PT psoriasis and as a vaccine against bacterial, fungal, viral or parasitic  
 PT infection.

XX PS Disclosure; Fig 4; 124pp; English.

XX CC The present invention describes a chimpanzee C68 adenoviral capsid  
 CC protein, which is substantially free of other viral proteins with which  
 CC it is naturally associated, comprising: (a) a hexon protein comprising  
 CC 513-amino acid sequence (see ABP56675); (b) a penton protein comprising  
 CC 534-amino acid sequence (see ABP56675); (c) a fibre protein comprising  
 CC 425-amino acid sequence (see ABP56675); or (d) a unique fragment of any  
 CC of (A)-(C) comprising 8-amino acid residues in length. Also described:  
 CC (1) a novel adenovirus serotype comprising a unique fragment of the C68  
 CC hexon protein fused to a heterologous adenovirus hexon peptide; (2) a  
 CC recombinant or pseudotyped adenovirus comprising a capsid of the novel  
 CC adenovirus serotype encapsidating a molecule for delivery to a target  
 CC cell; (3) an adenoviral capsid; (4) a pharmaceutical composition; (5)  
 CC rapid screening of recombinant constructs; and (6) a host cell. The  
 CC chimpanzee C68 adenoviral capsid protein has cytostatic, antiparasitic

C and antimicrobial activities, and can be used in vaccines. The chimpanzee  
 C C68 adenoviral capsid protein is useful for preparing a composition for  
 C treating hyperproliferative conditions e.g., cancer or psoriasis and as a  
 C vaccine against bacterial, fungal, viral or parasitic infection. The  
 C present sequence represents a human adenovirus capsid hexon protein which  
 C is given in comparison with the chimpanzee C68 adenovirus hexon protein  
 C in the exemplification of the present invention

X X Sequence 338 AA;

Query Match 86.8%; Score 165; DB 6; Length 338;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-15;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 5 GVINTETLTAKVKPTQENGWEKDATEFSK 35  
 |||||  
 b 295 GVINTETLTAKVKPTQENGWEKDATEFSK 325

RESULT 4  
 AAW79539  
 D AAW79539 standard; protein; 952 AA.  
 X C AAW79539;

17-OCT-2003 (revised)  
 11-JAN-1999 (first entry)

E Adenovirus serotype 5 hexon protein.

X X Adenovirus serotype 5; Ad5; hexon; coat protein; gene therapy; vector.  
 X M Adenovirus serotype 5.  
 X S Human adenovirus type 5.

H Key Location/Qualifiers  
 T Region 132..320  
 T /label= L1\_loop  
 T /note= "Claim 7"  
 T Region 138..182  
 T /label= HVR1 region  
 T /note= "Claim 7"  
 T Region 188..194  
 T /label= HVR2 region  
 T /note= "Claim 7"  
 T Region 212..219  
 T /label= HVR3 region  
 T /note= "Claim 7"  
 T Region 248..261  
 T /label= HVR4 region  
 T /note= "Claim 7"  
 T Region 263..283  
 T /label= HVR5 region  
 T /note= "Claim 7"  
 T Region 305..316  
 T /label= HVR6 region  
 T /note= "Claim 7"  
 T Region 412..462  
 T /label= L2 loop  
 T /note= "Claim 7"  
 T Region 422..450  
 T /label= HVR7 region  
 T /note= "Claim 7"  
 T Misc-difference 951  
 T /label= Gln, His, Thr  
 T Misc-difference 952  
 T /label= Gln, His, Thr

WO9840509-A1.

17-SEP-1998.

13-MAR-1998; 98WO-US005033.

XX

PR 13-MAR-1997; 97US-00816346.  
 XX (CORR ) CORNELL RES FOUND INC.  
 PA (GENV-) GENVEC INC.  
 XX Crystal RG, Falck-Pedersen E, Gall J, Kovesdi I, Wickham TJ;  
 PI WPI; 1998-506738/43.  
 XX DR N-PSDB; AAV61501.  
 XX Chimeric adenovirus coat protein - useful in, e.g. vector for gene  
 PT transfer to treat inherited genetic diseases.  
 XX Disclosure; Page 66-70; 112pp; English.

XX This is the amino acid sequence of the wild-type hexon protein of  
 CC adenovirus serotype 5 (Ad5). The invention provides a chimeric adenoviral  
 CC coat protein, particularly a chimeric adenovirus hexon protein, that has  
 CC a decreased ability or inability to be recognised by a neutralising  
 CC antibody directed against the corresponding wild-type adenovirus coat  
 CC protein. The chimeric adenoviral coat protein has a non-native amino acid  
 CC sequence, especially comprising a deletion of one or more regions of the  
 CC hexon protein, particularly a hypervariable region (HVR) of loop 1 or 2  
 CC of the hexon protein. DNA sequences (see AAW61502-23) encoding claimed  
 CC Ad2 and Ad5 chimeric coat proteins (see AAW79540-61) are claimed. Also  
 CC claimed are an adenovirus vector that comprises the chimeric adenovirus  
 CC coat protein, a method of genetically modifying a cell by contacting with  
 CC the vector, and a host cell that comprises the chimeric adenovirus coat  
 CC protein. The vector can be used for gene transfer, for the treatment of  
 CC inherited diseases (e.g. by carrying the cystic fibrosis (CF)  
 CC transmembrane conductance regulator cDNA to treat CF). It can also be  
 CC used to render certain cells susceptible to the killing action of certain  
 CC drugs, or to study the effects of expression of specific genes in a given  
 CC cell or tissue in vitro or in vivo. (Updated on 17-OCT-2003 to  
 CC standardise OS field)

SQ Sequence 952 AA;

Query Match 86.8%; Score 165; DB 2; Length 952;  
 Best Local Similarity 100.0%; Pred. No. 9.7e-15;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GVINTETLTAKVKPTQENGWEKDATEFSK 35  
 |||||  
 Db 419 GVINTETLTAKVKPTQENGWEKDATEFSK 449

RESULT 5  
 AAW63118  
 ID AAW63118 standard; protein; 952 AA.  
 XX AC AAW63118;

17-OCT-2003 (revised)  
 24-NOV-1998 (first entry)

Human adenovirus 5 hexon.  
 XX Adenovirus 5; Ad5; hexon; vector; gene therapy; vaccine.  
 XX Human adenovirus type 5.

XX Key Location/Qualifiers  
 FT Misc-difference 88  
 FT /note= "encoded by GGG"  
 FT Region 142..322  
 FT /label= L1  
 FT /note= "loop 1"  
 FT Region 405..517  
 FT /label= L2  
 FT /note= "loop 2"  
 FT Region 502..558  
 FT /label= L4

FT Misc-difference 602 /note= "loop 4"  
 FT FT /note= "encoded by ACG"

XX WO9832842-A1.

XX 30-JUL-1998.

XX 22-JAN-1998; 98WO-US001113.

XX 24-JAN-1997; 97US-00788674.

XX (GENE-) GENETIC THERAPY INC.

XX ROY S;

XX WPI; 1998-427938/36.

XX N-PSDB; AAV42656.

XX Modified adenoviral vaccine - comprises adenovirus with a hexon loop region of a second viral serotype.

XX Disclosure; Fig 1Aa-e; 67pp; English.

XX This is the amino acid sequence of the adenovirus 5 hexon. A modified adenovirus, in which the original adenovirus is of a first serotype within a first subgenus, and in which, in the modified version, at least a portion of at least one loop region of the hexon is removed and replaced with at least a portion of at least one loop region of the hexon of an adenovirus of a second serotype within a second subgenus, is new. Also claimed is a method of providing a therapeutic effect in a host comprising the administration of the adenovirus. The invention can be used as a vaccine against adenoviruses or as a gene therapy vector. The adenovirus 12 hexon amino acid sequence is also provided (see AAW63117). (Updated on 17-OCT-2003 to standardise OS field)

XX Sequence 952 AA;

Query Match 86.8%; Score 165; DB 2; Length 952;  
 Best Local Similarity 100.0%; Pred. No. 9.7e-15;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GVINTETLTKKVKTQENGWEKDATEFSDK 35

Db 419 GVINTETLTKKVKTQENGWEKDATEFSDK 449

RESULT 6

AAW79559

ID AAW79559 standard; peptide; 29 AA.

XX AAW79559;

DT 17-OCT-2003 (revised)

DT 11-JAN-1999 (first entry)

XX Adenovirus serotype 5 hypervariable region HVR7.

XX Adenovirus serotype 5; Ad5; hexon; coat protein; gene therapy; vector.

XX Human adenovirus type 5.

XX WO9840509-A1.

XX 17-SEP-1998.

XX 13-MAR-1998; 98WO-US005033.

XX 13-MAR-1997; 97US-00816346.

XX (CORR ) CORNELL RES FOUND INC.

XX (GENV-) GENVEC INC.

XX

PI

XX

DR WPI; 1998-506738/43.

XX N-PSDB; AAV61521.

XX

PT Chimeric adenovirus coat protein - useful in, e.g. vector for gene transfer to treat inherited genetic diseases.

XX

PS Claim 7; Page 88; 112pp; English.

XX

XX This is the amino acid sequence of hypervariable region HVR7 of the hexon protein (see AAW9559) of adenovirus serotype 5 (Ad5). It is encoded by a claimed DNA sequence (see AAV61521). The invention provides a chimeric adenoviral coat protein, particularly a chimeric adenovirus hexon protein, that has a decreased ability or inability to be recognised by a neutralising antibody directed against the corresponding wild-type adenovirus coat protein. The chimeric adenoviral coat protein has a non-native amino acid sequence, especially comprising a deletion of an internal hexon protein sequence, preferably a hypervariable region or entire loop. DNA sequences (see AAV61502-23) encoding claimed Ad2 and Ad5 chimeric coat proteins (see AAW9540-61) are claimed. Also claimed are an adenovirus vector that comprises the chimeric adenovirus coat protein, a method of genetically modifying a cell by contacting it with the vector, and a host cell that comprises the chimeric adenovirus coat protein. The vector can be used for gene transfer, for the treatment of inherited diseases. It can also be used to render certain cells susceptible to the killing action of certain drugs, or to study the effects of expression of specific genes in a given cell or tissue in vitro or in vivo. (Updated on 17-OCT-2003 to standardise OS field)

XX Sequence 29 AA;

Query Match 79.5%; Score 151; DB 2; Length 29;

Best Local Similarity 100.0%; Pred. No. 1.8e-14;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 NTETLTKKVKTQENGWEKDATEFSDK 35

Db 1 NTETLTKKVKTQENGWEKDATEFSDK 28

RESULT 7

ABB79316

ID ABB79316 standard; protein; 93 AA.

XX ABB79316;

XX 12-AUG-2002 (first entry)

XX Human ovary specific protein SEQ ID NO:113.

XX Human; ovary specific nucleic acid; OSNA; ovary specific protein; OSP;  
 XX ovary specific gene; OSG; ovarian cancer; immune response; metastasis.

XX Homo sapiens.

XX WO200240535-A2.

XX

XX 23-MAY-2002.

XX

XX 20-NOV-2001; 2001WO-US045011.

XX

XX 20-NOV-2000; 2000US-0252061P.

XX

XX 27-NOV-2000; 2000US-0253257P.

XX

XX (DIAD-) DIADEXUS INC.

XX

XX Salceda S, Macina RA, Recipon H, Caferkey R, Sun Y, Liu C;

XX

XX WPI; 2002-471617/50.

XX

XX New ovary specific genes and proteins, useful as a vaccine for treating

XX patients with ovarian cancer, or for diagnosing and monitoring the

PT

CC expression of bacterial genes, and for antibacterial screening. The  
 CC protein can be also used as a vaccine  
 XX  
 SQ Sequence 151 AA;  
 Query Match 29.5%; Score 56; DB 2; Length 151;  
 Best Local Similarity 32.4%; Pred. No. 7.1;  
 Matches 12; Conservative 8; Mismatches 13; Indels 4; Gaps 1;  
 QY 2 KGKGVINTETLTK-----VKPKTQENGWEXKDATEFSD 34  
 DB 86 EGRRVFTVEVDSVQPLEPKNAQNGGQKQNEFQD 122  
 RESULT 9  
 AAU37429  
 ID AAU37429 standard; protein; 167 AA.  
 XX  
 AC AAU37429;  
 XX 14-FEB-2002 (first entry)  
 DE Staphylococcus aureus cellular proliferation protein #1599.  
 KW Antisense; prokaryotic cellular proliferation protein; antibiotic;  
 KW antibacterial; drug design.  
 XX Staphylococcus aureus.  
 OS  
 FN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 XX 21-MAR-2001; 2001WO-US009180.  
 XX  
 PR 21-MAR-2000; 2000US-0191078P.  
 PR 23-MAY-2000; 2000US-0206848P.  
 PR 26-MAY-2000; 2000US-0207727P.  
 PR 23-OCT-2000; 2000US-0242578P.  
 PR 27-NOV-2000; 2000US-0253625P.  
 PR 22-DEC-2000; 2000US-0257931P.  
 PR 16-FEB-2001; 2001US-0269308P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX  
 DR WPI; 2001-611495/70.  
 DR N-PSDB; AAS55288.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids.  
 XX  
 PS Example 3; SEQ ID NO 13022; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the genes,  
 CC their use in the discovery of novel antibiotics, the essential genes  
 CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
 CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
 CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
 CC useful for the identification of potential new targets for antibiotic  
 CC development. The antisense nucleic acids can also be used to identify  
 CC proteins used in proliferation, to express these proteins, and to obtain  
 CC antibodies capable of binding to the expressed proteins. The proteins can  
 CC be used to screen compounds in rational drug discovery programmes. The  
 CC antisense nucleic acid sequence is also useful to screen for homologous  
 CC nucleic acids which are required for cell proliferation in a wide variety  
 CC of organisms. The present sequence represents an essential prokaryotic  
 CC cellular proliferation protein. Note: The sequence data for this patent  
 CC did not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at



Gaps 1;





Search completed: March 7, 2004, 13:35:12  
Job time : 57.3846 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

DM protein - protein search, using sw model

Run on: March 7, 2004, 13:28:01 ; Search time 8.84615 Seconds  
(without alignments)  
206.017 Million cell updates/sec

Title: US-09-643-458B-15  
Perfect score: 190  
Sequence: 1 CKGKGVINTETLTKVVKPTQENGWEKDATEFSDK 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	165	86.8	951	1 HEX_ADE05	P04133 human adeno
2	59	31.1	944	1 HEX_ADE03	P36849 human adeno
3	56	29.5	323	1 PRG1_CABEL	O18282 caenorhabdi
4	56	29.5	1338	1 ACIN_MOUSE	Q9J1X8 mus musculu
5	56	29.5	1341	1 ACIN_MOUSE	Q9J1X8 mus musculu
6	54	28.4	317	1 FECR_ECOLI	P23485 escherichia
7	54	28.4	412	1 AFKB_ARATH	P46573 arabidopsis
8	53	27.9	110	1 CLS1_BRAJA	Q99JW5 bradyrhizob
9	53	27.9	677	1 YD64_MYCPN	P75417 mycoplasma
10	52	27.4	502	1 YDGA_ECOLI	P77804 escherichia
11	51.5	27.1	447	1 HEX_ADE04	P36850 human adeno
12	51.5	27.1	1267	1 VU3_REOVL	P17376 reovirus (c
13	51	26.8	1905	1 Y659_PASMU	Q9cmz1 pasteurella
14	50.5	26.6	382	1 DACB_BACSU	P35150 bacillus su
15	50.5	26.6	751	1 VP4_ROTGI	P15155 rotavirus (
16	50	26.3	376	1 DNAJ_METTH	O27352 methanobact
17	50	26.3	410	1 APKA_ARATH	Q06548 arabidopsis
18	49	25.8	707	1 TRP_YEAS	P00931 saccharomyc
19	48.5	25.5	309	1 YNR1_YEAS	P33851 saccharomyc
20	48.5	25.5	471	1 E1L4_ARATH	Q91X16 arabidopsis
21	48	25.3	187	1 DDF1_YEAS	Q99321 saccharomyc
22	48	25.3	381	1 DRPH_BACSH	P23307 bacillus sp
23	48	25.3	452	1 AMT_CORGL	P54146 corynebacte
24	48	25.3	932	1 SCF3_SYNY3	Q55709 synecocyst
25	48	25.3	2009	1 SEC7_YEAS	P11075 saccharomyc
26	48	25.3	3224	1 RBF2_HUMAN	P49792 homo sapien
27	47.5	25.0	397	1 PGLR_BRANA	P35337 brassica na
28	47.5	25.0	565	1 YOG1_CAEEL	P34610 caenorhabd
29	47.5	25.0	1091	1 C1C2_HUMAN	P54289 homo sapien
30	47.5	25.0	1091	1 C1C2_RAT	P54290 rattus norv
31	47.5	25.0	1106	1 C1C2_RABIT	P13806 oryctolagus
32	47.5	25.0	1634	1 DPOL_METFA	Q58295 methanococ
33	47	24.7	209	1 R55_METMA	Q8tr87 methanosarc

34	47	24.7	209	1 R55_METMA	Q8pv30 methanosarc
35	47	24.7	226	1 R55_METMA	Q8tza6 methanopyru
36	47	24.7	397	1 DNJ1_ALLPO	Q03363 allium porr
37	47	24.7	418	1 DNJ2_ALLPO	P42824 allium porr
38	47	24.7	486	1 YNSU_YEAST	P53742 saccharomyc
39	47	24.7	528	1 V58K_BSMV	P04867 barley stri
40	47	24.7	1001	1 IF2_SYNY3	P72689 synecocyst
41	47	24.7	1024	1 RIP3_MOUSE	P97434 mus musculu
42	47	24.7	1029	1 RIP3_RAT	Q9ere6 rattus norv
43	46.5	24.5	183	1 SAV1_STRVL	Q53533 streptomyce
44	46.5	24.5	183	1 SAV2_STRVL	Q53533 streptomyce
45	46.5	24.5	183	1 SAV_STRAV	P22629 streptomyce

ALIGNMENTS

RESULT 1

HEX\_ADE05  
ID HEX\_ADE05 STANDARD; PRT; 951 AA.  
AC P04133;  
DT 01-NOV-1986 (Rel. 03, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hexon protein (Late protein 2).  
GN P11.  
OS Human adenovirus type 5.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
OX NCBI\_TaxID=28285;  
RN [1]\_TaxID=28285;  
RP SEQUENCE FROM N.A.  
RA MEDLINE=84212465; PubMed=6202684;  
RX Kinloch R., Mackay N., Mautner V.;  
RT "Adenovirus hexon. Sequence comparison of subgroup C serotypes 2 and 5."  
RL J. Biol. Chem. 259:6431-6436(1984).  
RN [2]  
RP COMPLETE GENOME.  
RA MEDLINE=92087470; PubMed=1727603;  
RX Chroboczek J., Bieber F., Jacrot B.;  
RT "The sequence of the genome of adenovirus type 5 and its comparison with the genome of adenovirus type 2."  
RL Virology 186:280-285(1992).  
CC -!- FUNCTION: This protein is one of the structural proteins in the viral coat and is synthesized during late infection.  
CC -!- SUBUNIT: Homotrimer (By similarity).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: J01966; AAD15293.1; -;  
CC EMBL: M73260; -; NOT ANNOTATED\_CDS.  
CC EMBL: X02957; CAA26753.1; -;  
CC FIR: A03849; EXAD5.  
CC PDB: 1P30; 11-NOV-03.  
CC InterPro: IPR000736; Adeno hexon.  
CC Pfam: PF01065; Adeno hexon; 1.  
CC Pfam: PF03678; Adeno hexon C; 1.  
CC ProDom: PD002815; Adeno hexon; 1.  
CC Coat protein; Hexon protein; Late protein; 3D-structure.  
FT INIT MET 0 0 BY SIMILARITY.  
FT DOMAIN 133 158 ASP/GLU-RICH (ACIDIC)  
SQ SEQUENCE 951 AA; 107875 MW; EF0F9A24961B117F CRC64;

Query Match 86.8%; Score 165; DB 1; Length 951;  
Best Local Similarity 100.0%; Pred. No. 3.4e-14;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GVINTETLTQVKTGOENGWEKDATEFSK 35  
 DB 418 GVINTETLTQVKTGOENGWEKDATEFSK 448

## RESULT 2

HEX\_ADE03  
 ID HEX\_ADE03 STANDARD; PRT; 944 AA.  
 AC P36849;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Hexon protein (Late protein 2).  
 GN P11  
 OS Human adenovirus type 3.  
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
 OX NCBI\_TaxID=45659;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Isolate GB;  
 RX MEDLINE=95407102; PubMed=7676636;  
 RA Ping-Akerblom P., Trijssenaar J., Adrian T.;  
 RT "Sequence characterization and comparison of human adenovirus  
 RT subgenus B and E hexons.";  
 RL Virology 212:232-236(1995).  
 RN [2]  
 RP REVISIONS.  
 RA Ping-Akerblom P.;  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: This protein is one of the structural proteins in the  
 CC viral coat and is synthesized during late infection.  
 CC -!- SUBUNIT: Homotrimer (By similarity).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

Query Match 31.1%; Score 59; DB 1; Length 944;  
 Best Local Similarity 48.0%; Pred. No. 5;  
 Matches 12; Conservative 2; Mismatches 11; Indels 0; Gaps 0;  
 QY 5 GVINTETLTQVKTGOENGWEKDA 29  
 DB 412 GIGPGHTYQGIKVTDDTNGWEKDA 436

## RESULT 3

FRG1 CAEEL  
 ID FRG1 CAEEL STANDARD; PRT; 323 AA.  
 AC O18282;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE FRG1 protein homolog.  
 GN ZK1010.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Gardner A.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO THE FRG1 FAMILY.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; Z82083; CAB04969.1; -;  
 DR PIR; T27640; T27640.  
 DR WomPep; ZK1010.3; CE15499.  
 DR InterPro; IPR008999; Actin\_crosslink.  
 KW Hypothetical protein.  
 SQ SEQUENCE 323 AA; 36331 MW; 4AADA219FF4D0 CRC64;

Query Match 29.5%; Score 56; DB 1; Length 323;  
 Best Local Similarity 38.2%; Pred. No. 4.3;  
 Matches 13; Conservative 3; Mismatches 18; Indels 0; Gaps 0;

QY 2 KGKGVINTETLTQVKTGOENGWEKDATEFSK 35  
 DB 77 KEKKKKNDKEKIDPDVTENGGRKIADEFDMK 110

## RESULT 4

ID ACIN MOUSE STANDARD; PRT; 1338 AA.  
 AC Q9JIX8; Q9CSN7; Q9CSR9; Q9CSX7; Q9R046; Q9R047;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Apoptotic chromatin condensation inducer in the nucleus (Acinus).  
 GN ACINUS.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).  
 RX MEDLINE=99418558; PubMed=10490026;  
 RA Sahara S., Aoto M., Eguchi Y., Imamoto N., Yoneda Y., Tsujimoto Y.;  
 RT "Acinus is a caspase-3-activated protein required for apoptotic  
 RT chromatin condensation.";  
 RL Nature 401:168-173(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RP Mamoru A., Setuko S., Yoshihide T.;  
 RA "Molecular cloning of murine acinusL, a gene for apoptotic chromatin  
 RT condensation.";  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-1190 FROM N.A. (ISOFORMS 1; 2 AND 4).  
 RC STRAIN=C57BL/6J; TISSUE=Embryo, and Pancreas;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer J.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,



```
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1; Synonyms=L;
CC IsoId=Q9UKV3-1; Sequence=Displayed;
CC Name=2; Synonyms=S;
CC IsoId=Q9UKV3-2; Sequence=VSP_004025, VSP_004028;
CC Name=3; Synonyms=S;
CC IsoId=Q9UKV3-3; Sequence=VSP_004026, VSP_004029;
CC Name=4;
CC IsoId=Q9UKV3-4; Sequence=VSP_004027;
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- PTM: Undergoes proteolytic cleavage; the processed form is active,
CC contrary to the uncleaved form.
CC -!- SIMILARITY: Contains 1 SAP domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF124726; AAD56724.1; -
CC EMBL; AF124727; AAD56725.1; -
CC EMBL; AF124728; AAD56726.1; -
CC EMBL; AL050382; CAB43681.1; -
CC EMBL; BX247975; CAD62309.1; -
CC EMBL; AB014570; BAA31645.2; -
CC Genew; HGNC:17066; ACINUS.
CC MIM; 604562; -
CC GO; GO:0005634; C:nucleus; IDA.
CC GO; GO:0016887; F:ATPase activity; NAS.
CC GO; GO:0019899; F:nucleic acid binding; NAS.
CC GO; GO:0003676; F:nucleic acid binding; NAS.
CC GO; GO:0030263; P:apoptotic chromosome condensation; IDA.
CC GO; GO:0030218; P:erythrocyte differentiation; IEP.
CC GO; GO:0045857; P:positive regulation of monocyte differentiation. . .; IEP.
CC InterPro; IPR003034; SAP.
CC Pfam; PF02037; SAP; 1.
CC SMART; SM00513; SAP; 1.
CC PROSITE; PS00800; SAP; 1.
CC Apoptosis; Nuclear protein; Alternative splicing.
CC DOMAIN 72 106
CC SAP.
CC GLU-RICH.
CC DOMAIN 142 442
CC SER-RICH.
CC DOMAIN 573 676
CC PRO-RICH.
CC DOMAIN 1114 1131
CC ARG/ASP/GLU/LYS-RICH.
CC DOMAIN 1132 1341
CC CLEAVAGE (BY CASPASE-3).
CC SITE 1093 1094
CC Missing (in isoform 2).
CC VARSPLIC 1 727
CC /FTid=VSP_004025.
CC Missing (in isoform 3).
CC VARSPLIC 1 758
CC /FTid=VSP_004026.
CC Missing (in isoform 4).
CC VARSPLIC 1 1152
CC /FTid=VSP_004027.
CC VARSPLIC 728 766
CC GSPKCEAEAEPPAATQPTSTQTSHTLPESEIRIHVV
CC -> MSPADRCRSANTTEPATTSLLALLLQDSRTGL
CC p (in isoform 2).
CC /FTid=VSP_004028.
CC SERIHVV -> MSESEK (in isoform 3).
CC /FTid=VSP_004029.
CC D->A: ABOLISHES CLEAVAGE BY CASP3 AND
CC CHROMATIN CONDENSATION ACTIVITY.
CC Q -> H (IN REF. 4).
CC CONFLICT 139 139
CC SEQUENCE 1341 AA; 151887 MW; 8FE286681F83AB5C CRC64;
CC -----
CC Query Match 29.5%; Score 56; DB 1; Length 1341;
CC Best Local Similarity 42.9%; Pred.No. 18;
CC Matches 12; Conservative 5; Mismatches 11; Indels 0; Gaps 0;
CC -----
CC QY 4 KGVNTETLTAKVKPKTGQENGWEDATE 31
CC DB 913 KGLKICRTVTQVPAEGQENGQREEE 940
```

```
RESULT 6
FECD_ECOLI
ID FECD_ECOLI STANDARD; PRT; 317 AA.
AC P23485;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein fecr.
GN FECD OR B4292.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
CX NCBI_TaxID=562;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=K12;
RX MEDLINE=91072220; PubMed=2254251;
RA van Hove B., Staudenmaier H., Braun V.;
RT "Novel two-component transmembrane transcription control: regulation
RT of iron dicitrate transport in Escherichia coli K-12.";
RL J. Bacteriol. 172:5749-5758 (1990).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119 (1995).
CC -!- FUNCTION: REGULATION OF IRON DICITRATE TRANSPORT. IN THE ABSENCE
CC OF CITRATE FECD INACTIVATES FECD. FECD IS PROBABLY A SENSOR THAT
CC RECOGNIZES IRON DICITRATE IN THE PERIPLASM.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M63115; AAA23767.1; -
CC EMBL; U14003; AAA97188.1; -
CC EMBL; AB000500; AAC77248.1; -
CC PIR; B37804; B37804.
CC EcoGene; EG10292; fecr.
CC InterPro; IPR006860; FecR.
CC Pfam; PF04773; FecR; 1.
CC Iron transport; Periplasmic; Sensory transduction; Complete proteome.
CC KW SEQUENCE 317 AA; 35532 MW; 798EC615815AF9D7 CRC64;
CC -----
CC Query Match 28.4%; Score 54; DB 1; Length 317;
CC Best Local Similarity 47.6%; Pred.No. 7;
CC Matches 10; Conservative 1; Mismatches 10; Indels 0; Gaps 0;
CC -----
CC QY 15 VKPKTGQENGWEDATEPESDK 35
CC DB 232 VKPLDDSTSTWKDILSFSDK 252
CC -----
RESULT 7
APKB_ARATH
ID APKB_ARATH STANDARD; PRT; 412 AA.
AC P46573; Q9SLH5;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein kinase APK1B (EC 2.7.1.-).
GN APK1B OR AT2G28930 OR T914.1.
```

OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 NCBI\_TaxID=3702;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.B., Feldblyum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,  
 RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,  
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,  
 RA Goodman H.M., Somerville C.R., Coppenhaver G.P., Preuss D.,  
 RA Venter J.C.;  
 RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 thaliana.";  
 RT Nature 402:761-768(1999).  
 RL [2]  
 RP SEQUENCE OF 143-346 FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=93081726; PubMed=1450380;  
 RA Hirayana T., Oka A.;  
 RA "Novel protein kinase of Arabidopsis thaliana (APK1) that  
 phosphorylates tyrosine, serine and threonine.";  
 RL Plant Mol. Biol. 20:653-662(1992).  
 CC -!- FUNCTION: Possible bi-functional kinase. In vitro, it exhibits  
 CC serine/threonine activity. In vivo, can phosphorylate tyrosine  
 CC residues of limited substrates (By similarity).  
 CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AC005315; AAC33221.1; --  
 CC EMBL; D10152; BAA20968.1; --  
 DR PIR; I02725; T02725.  
 DR InterPro; IPR000719; Prot kinase.  
 DR InterPro; IPR008271; Ser\_thr\_pkin\_AS.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1  
 DR PRINTS; PR00109; TYRKINASE  
 DR ProDom; PD000001; Prot.kinase; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR Transferrase; Serine/threonine-protein kinase; Tyrosine-protein kinase;  
 KW ATP-binding; Multigene family; Myristate; Lipoprotein.  
 FT LIPID 2 2 N-myristoyl glycine (By similarity).  
 FT DOMAIN 69 356 PROTEIN\_KINASE.  
 FT NP\_BIND 75 83 ATP (By similarity).  
 FT BINDING 107 107 ATP (By similarity).  
 FT ACT\_SITE 204 204 BY SIMILARITY.  
 SQ SEQUENCE 412 AA; 45746 MW; E81CA0B1A626A5DA CRC64;  
 Query Match 28.4%; Score 54; DB 1; Length 412;  
 Best Local Similarity 38.7%; Pred. No. 10;  
 Matches 12; Conservative 5; Mismatches 6; Indels 8; Gaps 1;  
 QY 4 KGVINTELTITKVKPTG-----QENGWE 26  
 Db 85 KGWIDEQLTASKECTGVIVAKLNQDGMQ 115

RESULT 8  
 CLS1\_BRAJA

ID CLS1\_BRAJA STANDARD; PRT; 110 AA.  
 AC Q89JW5;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE ATP-dependent Clp protease adaptor protein clps 1.  
 DE CLPS1 OR BJU5154.  
 GN GN  
 OS Bradyrhizobium japonicum.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Bradyrhizobiaceae; Bradyrhizobium.  
 NCBI\_TaxID=375;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=USDA 110;  
 RX MEDLINE=22484998; PubMed=12597275;  
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,  
 RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,  
 RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,  
 RA Tabata S.;  
 RA "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
 Bradyrhizobium japonicum USDA110.";  
 RT DNA Res. 9:189-197(2002).  
 CC -!- FUNCTION: Involved in the modulation of the specificity of the  
 CC clpAP-mediated ATP-dependent protein degradation (By similarity).  
 CC -!- SUBUNIT: Binds to the N-terminal domain of the chaperone ClpA (By  
 CC similarity).  
 CC -!- SIMILARITY: Belongs to the clpS family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF005954; BAC50419.1; ALT\_INIT.  
 DR HAMAP; MF\_00302; -; 1.  
 DR InterPro; IPR003769; DUF174.  
 DR Pfam; PF02617; DUF174; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 110 AA; 12637 MW; A55016BBF71E80BA CRC64;  
 Query Match 27.9%; Score 53; DB 1; Length 110;  
 Best Local Similarity 62.5%; Pred. No. 3.6;  
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 QY 8 NTETLTKKVKPTGOEN 23  
 Db 15 NTSVITKKVKPTKRN 30  
 RESULT 9  
 YD64\_MYCPN STANDARD; PRT; 677 AA.  
 ID\_YD64\_MYCPN  
 AC P75417.  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein MPN364 (H91\_orf677).  
 GN MPN364 OR MP472.  
 OS Mycoplasma pneumoniae.  
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 NCBI\_TaxID=2104;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=ATCC 29342 / M129;  
 RX MEDLINE=97105885; PubMed=8948633;  
 RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,  
 RA Hermann R.;  
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
 pneumoniae.";  
 RL Nucleic Acids Res. 24:4420-4449(1996).

```

CC -!- SIMILARITY: BELONGS TO THE MG185 / MG260 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB000047; AAB96120.1; -.
DR PIR; S73798; S73798.
DR InterPro; IPR004890; Lipoprotein_10.
DR InterPro; IPR004984; Lipoprotein_X.
DR Pfam; PF03202; Lipoprotein_10; 1.
DR Pfam; PF03305; Lipoprotein_X; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 677 AA; 75591 MW; 5785868BD679F04D CRC64;

Query Match 27.9%; Score 53; DB 1; Length 677;
Best Local Similarity 35.3%; Pred. No. 23;
Matches 12; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

QY 2 KKGVINTETLTQKPKTGOENGWEKDAEPFSDK 35
DB 116 KKGTEKTTTLTLTKTSSQNSQSTTKGKATSD 149

RESULT 10
YDGA_ECOLI
ID YDGA_ECOLI STANDARD; PRT; 502 AA.
AC P77804; Q47707;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hypothetical protein ydga.
GN ydga OR B1614.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA Liang W.J., Wilson K.J., Jefferson R.A.;
RA Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Blattner F.R., Plunkett G. III, Bloch C.A., Berna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RA Science 277:1453-1474 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12.
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishio Y., Oshima T., Saito N.,
RA Sampei G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RA "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RA corresponding to the 28.0-40.1 min region on the linkage map.";
RA DNA Res. 3:363-377 (1996).
CC -!- SIMILARITY: TO E.COLI YIHF AND H.INFLUENZAE H1236.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 422.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M14641; AAA68926.1; AUT_FRAME.
DR EMBL; AB000257; AAC74686.1; -.
DR EMBL; D90805; BAA15366.1; -.
DR EMBL; D90804; BAA15362.1; -.
DR PIR; H64917; H64917.
DR EcoGene; EG12457; ydga.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 5 25 POTENTIAL.
FT CONFLICT 186 198 GKALISLGEAQSG -> AKPSFPFGRKVV (IN REF. 1).
FT CONFLICT 322 323 LA -> SR (IN REF. 1).
SQ SEQUENCE 502 AA; 54689 MW; D061576FA5DBAC4C CRC64;

Query Match 27.4%; Score 52; DB 1; Length 502;
Best Local Similarity 41.7%; Pred. No. 23;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 4 KGVINTETLTQKPKTGOENGWEK 27
DB 62 RGVFSQQLLVKPIAKENPWIK 85

RESULT 11
HEX_ADE04
ID HEX_ADE04 STANDARD; PRT; 447 AA.
AC P36850;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hexon protein (Late protein 2) (Fragment).
GN pII.
OS Human adenovirus type 4.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=28280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Isolate RJ-67;
RX MEDLINE=95407102; PubMed=7676636;
RA Pring-Akerblom P., Trijssenar J., Adrian T.;
RA "Sequence characterization and comparison of human adenovirus
RA subgenus B and E hexons.";
RA Virology 212:232-236 (1995).
CC -!- FUNCTION: This protein is one of the structural proteins in the
CC viral coat and is synthesized during late infection.
CC -!- SUBUNIT: Homotrimer (By similarity).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X76550; CAA54052.1; -.
DR PIR; S39296; S39296.
DR HSP; P03277; IDHX.
DR InterPro; IPR000736; Adeno_hexon.
DR Pfam; PF01065; Adeno_hexon; 1.
DR ProDom; PD02815; Adeno_hexon; 1.
KW Coat protein; Hexon protein; Late protein.
FT NON TER 1 1
FT NON TER 447 447
SQ SEQUENCE 447 AA; 49553 MW; A7AE1977F707BD4D CRC64;

Query Match 27.1%; Score 51.5; DB 1; Length 447;

```



Best Local Similarity 40.6%; Pred. No. 24;  
Matches 13; Conservative 4; Mismatches 14; Indels 1; Gaps 1;

QY 3 GKGVINTETLTQVKTGNGWEXKDATEFSD 34  
DB 301 GVLGTDYQGVKVTGADSEK-WDKDDTTVSN 331

RESULT 12  
VL3\_RECVL STANDARD; PRT; 1267 AA.  
AC P17376;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-NOV-1990 (Rel. 16, Last annotation update)  
DE Minor core protein lambda 3.  
GN L1.  
OS Reovirus (type 1 / strain Lang).  
OC Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.  
OX NCBI\_TaxID=10884;  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89163254; PubMed=2922925;  
RA Wiener J.R., Joklik W.K.;  
RT "The sequences of the reovirus serotype 1, 2, and 3 L1 genome  
RT segments and analysis of the mode of divergence of the reovirus  
RT serotypes.";  
RL Virology 169:194-203(1989).  
CC -!- MISCELLANEOUS: IT IS UNLIKELY THAT CORES CONTAIN MORE THAN 12  
CC -!- MOLECULES OF PROTEIN LAMBDA 3.  
CC -!- SIMILARITY: VERY HIGH WITH OTHER REOVIRUS TYPES.  
DR PIR; A30121; MWK31.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007097; RNA\_pol\_rec.  
KW Core protein.  
SQ SEQUENCE 1267 AA; 142354 MW; 157BFD4D664FCDB CRC64;

Query Match 27.1%; Score 51.5; DB 1; Length 1267;  
Best Local Similarity 40.7%; Pred. No. 68;  
Matches 11; Conservative 6; Mismatches 7; Indels 3; Gaps 1;

QY 5 GVINTELTQVKTGNGWEXKDATEFSD 28  
DB 746 GKVNSITQVLSKYEGRGKID 772

RESULT 13  
Y659\_PASMU STANDARD; PRT; 1905 AA.  
AC Q9CMZ1;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hypothetical UPF0192 protein PM0659 precursor.  
GN PM0659.  
OS Pasteurella multocida.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OX Pasteurellaceae; Pasteurella.  
OX NCBI\_TaxID=747;  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=PM70;  
RX MEDLINE=21145856; PubMed=11248100;  
RA May B., Zhang Q., Li L.S., Paustian M.L., Whittam T.S., Kapur V.;  
RT "Complete genomic sequence of Pasteurella multocida PM70.";  
RT Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
CC -!- SIMILARITY: Belongs to the UPF0192 family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AR006102; AAK02743.1;  
DR InterPro; IPR00437; Prok\_lipoprot\_S.  
KW Hypothetical protein; Signal; Complete proteome.  
FT SIGNAL 1 16  
FT CHAIN 17 1905 HYPOTHETICAL PROTEIN PM0659.  
SQ SEQUENCE 1905 AA; 214427 MW; FED71CEBD61F7C78 CRC64;

Query Match 26.8%; Score 51; DB 1; Length 1905;  
Best Local Similarity 47.8%; Pred. No. 1,2e+02;  
Matches 11; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 2 KKGVINTETLTQVKTGNGWEXKDATEFSD 24  
DB 84 KGVNSITQVLSKYEGRGKID 106

RESULT 14  
DABC\_BACSU STANDARD; PRT; 382 AA.  
AC P35150;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Penicillin-binding protein 5\* precursor (D-alanyl-D-alanine  
DE carboxypeptidase) (EC 3.4.16.4) (DD-peptidase) (DD-carboxypeptidase)  
DE (PBP-5\*).  
GN DABC OR BSU23190.  
OC Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN  
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
RX MEDLINE=92193254; PubMed=1548223;  
RA Buchanan C.E., Ling M.-L.;  
RT "Isolation and sequence analysis of dacB, which encodes a  
RT sporulation-specific penicillin-binding protein in Bacillus  
RT subtilis.";  
RL J. Bacteriol. 174:1717-1725(1992).  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168 / Marburg;  
RX MEDLINE=95020538; PubMed=7934829;  
RA Sorokin A.V., Zumstein E., Azevedo V., Ehrlich S.D., Serror P.;  
RT "The organization of the Bacillus subtilis 168 chromosome region  
RT between the spoVA and serA genetic loci, based on sequence data.";  
RL Mol. Microbiol. 10:385-395(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Broutlet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Duetzsch A., Ehrlich S.D., Emerson P.T.,  
RA Eschian K.D., Ezringon J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Glim S.T., Glaser P., Goffeau A., Goughly E.J., Grandi G.,  
RA Giuseppe G., Guy B.J., Hsaga K., Hsaga J., Harwood C.R., Henaut A.,  
RA Hilbert H., Hoispey S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M.,  
RA Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y,



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2004, 13:31:21 ; Search time 38.0769 Seconds  
(without alignments)  
290.022 Million cell updates/sec

Title: US-09-643-458b-15

Perfect score: 190

Sequence: 1 CKRGVINTLTQVKPKTGQENGWEKDATEFSK 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 25.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	165	86.8	952	12 Q80RJ2	Q80RJ2 human adeno
2	165	86.8	952	12 Q805J1	Q805J1 human adeno
3	151	79.5	952	12 Q80R15	Q80R15 human adeno
4	143	75.3	52	12 Q9YIS1	Q9YIS1 human adeno
5	143	75.3	952	12 Q80R10	Q80R10 human adeno
6	64	33.7	226	10 Q93XC4	Q93XC4 brassica ol
7	62	32.6	226	10 Q93V78	Q93V78 brassica ol
8	61	32.1	944	12 Q80RH8	Q80RH8 human adeno
9	61	32.1	944	12 Q80RH6	Q80RH6 human adeno
10	59	31.1	344	12 Q91PN4	Q91PN4 human adeno
11	59	31.1	493	12 Q6S296	Q6S296 human adeno
12	58.5	30.8	264	17 Q8ZYL4	Q8ZYL4 pyrobaculum
13	58	30.5	1627	10 Q8LQJ0	Q8LQJ0 cryza sativ
14	57	30.0	265	16 Q98EE6	Q98EE6 rhizobium l
15	57	30.0	344	12 Q91PL8	Q91PL8 human adeno
16	57	30.0	344	12 Q91PN1	Q91PN1 human adeno

17	57	30.0	344	12 Q91PL9	Q91PL9 human adeno
18	57	30.0	344	12 Q91PM4	Q91PM4 human adeno
19	57	30.0	344	12 Q91PN3	Q91PN3 human adeno
20	57	30.0	344	12 Q91PM0	Q91PM0 human adeno
21	57	30.0	344	12 Q91PM8	Q91PM8 human adeno
22	57	30.0	344	12 Q91PM9	Q91PM9 human adeno
23	57	30.0	344	12 Q91PN0	Q91PN0 human adeno
24	57	30.0	344	12 Q91PM6	Q91PM6 human adeno
25	57	30.0	344	12 Q91PL6	Q91PL6 human adeno
26	57	30.0	344	12 Q91PL5	Q91PL5 human adeno
27	57	30.0	344	12 Q91PM7	Q91PM7 human adeno
28	57	30.0	344	12 Q91PM2	Q91PM2 human adeno
29	57	30.0	344	12 Q91PM1	Q91PM1 human adeno
30	57	30.0	344	12 Q91PL7	Q91PL7 human adeno
31	57	30.0	344	12 Q91PM5	Q91PM5 human adeno
32	57	30.0	344	12 Q91PM2	Q91PM2 human adeno
33	57	30.0	344	12 Q91PM3	Q91PM3 human adeno
34	57	30.0	944	12 Q80RJ7	Q80RJ7 human adeno
35	57	30.0	944	12 Q80RJ6	Q80RJ6 human adeno
36	57	30.0	944	12 Q80RJ5	Q80RJ5 human adeno
37	57	30.0	944	12 Q80RJ4	Q80RJ4 human adeno
38	57	30.0	944	12 Q80RJ3	Q80RJ3 human adeno
39	57	30.0	944	12 Q80RJ1	Q80RJ1 human adeno
40	57	30.0	944	12 Q80RJ0	Q80RJ0 human adeno
41	57	30.0	944	12 Q80RI9	Q80RI9 human adeno
42	57	30.0	944	12 Q80RI8	Q80RI8 human adeno
43	57	30.0	944	12 Q80RI7	Q80RI7 human adeno
44	57	30.0	944	12 Q80RI6	Q80RI6 human adeno
45	57	30.0	944	12 Q80RI4	Q80RI4 human adeno

#### ALIGNMENTS

RESULT 1  
Q80RJ2 PRELIMINARY; PRT; 952 AA.  
AC Q80RJ2;  
DT 01-JUN-2003 (TREMREL. 24, Created)  
DT 01-JUN-2003 (TREMREL. 24, Last sequence update)  
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)  
DE Hexon.  
OS Human adenovirus type 5.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
OX NCBI\_TaxID=28285;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KNH Ad99/5;  
RA Na B.-K., Lee J.-Y., Kim J.-H., Kim W.-J., Kang C.;  
RT "Sequence analysis of hexon protein of adenoviruses isolated in Korea."  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF542109; AAC24084.1; -;  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR000736; Adeno\_hexon.  
DR Pfam; PF01065; Adeno\_hexon\_1.  
DR Pfam; PF03678; Adeno\_hexon\_C; 1.  
DR ProDom; PD002815; Adeno\_hexon; 1.  
SQ SEQUENCE 952 AA; 107938 MW; E95940D564209EAB CRC64;

Query Match 86.8%; Score 165; DB 12; Length 952;  
Best Local Similarity 100.0%; Pred. No. 1.8e-13;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GVINTETLTQVKPKTGQENGWEKDATEFSK 35  
DB 419 GVINTETLTQVKPKTGQENGWEKDATEFSK 449

RESULT 2  
Q805J1 PRELIMINARY; PRT; 952 AA.  
ID Q805J1

```
AC Q805J1;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hexon.
OS Human adenovirus type 5.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=28285;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KNH Ad00/12, KNIH Ad00/19, and KNIH Ad01/1;
RA Na B.-K., Lee J.-Y., Kim J.-H., Kim W.-J., Kang C.;
RT "Sequence analysis of hexon protein of adenoviruses isolated in Korea.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF5421124; AAO24099.1; -.
DR EMBL; AF5421128; AAO24103.1; -.
DR EMBL; AF5421130; AAO24105.1; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000736; Adeno_hexon.
DR Pfam; PF01065; Adeno_hexon; 1.
DR ProDom; PD002815; Adeno_hexon; 1.
SQ SEQUENCE 952 AA; 107915 MW; C92738D3C0383802 CRC64;

Query Match 86.8%; Score 165; DB 12; Length 952;
Best Local Similarity 100.0%; Pred. No. 1.8e-13;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GVINTETLTVKPKTGOENGWEKDATEFSDK 35
DB 419 GVINTETLTVKPKTGOENGWEKDATEFSDK 449

RESULT 3
Q80RI5
ID Q80RI5 PRELIMINARY; PRT; 952 AA.
AC Q80RI5;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hexon.
OS Human adenovirus type 5.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=28285;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KNH Ad99/12;
RA Na B.-K., Lee J.-Y., Kim J.-H., Kim W.-J., Kang C.;
RT "Sequence analysis of hexon protein of adenoviruses isolated in Korea.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF5421116; AAO24091.1; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000736; Adeno_hexon.
DR Pfam; PF01065; Adeno_hexon; 1.
DR ProDom; PD002815; Adeno_hexon; 1.
SQ SEQUENCE 952 AA; 107915 MW; 0514E635D6CADA54 CRC64;

Query Match 79.5%; Score 151; DB 12; Length 952;
Best Local Similarity 87.1%; Pred. No. 1.4e-11;
Matches 27; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 GVINTETLTVKPKTGOENGWEKDATEFSDK 35
DB 419 GVINTETLTVKPKTGOENGWEKDATEFSDK 449

RESULT 4
Q9YLS1
```

```
ID Q9YLS1 PRELIMINARY; PRT; 52 AA.
AC Q9YLS1;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Hexon protein (Fragment).
GN HEXON.
OS Human adenovirus type 5.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=28285;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=95355;
RX MEDLINE=99143271; PubMed=986873;
RA Li O.G., Henningsen A., Juto P., Elgh P., Wadell G.;
RT "Use of restriction fragment analysis and sequencing of a serotype-specific region to type adenovirus isolates.";
RL J. Clin. Microbiol. 37:844-847(1999).
DR EMBL; AF051949; AAD18129.1; -.
DR HSP; P03277; IDHX.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000736; Adeno_hexon.
DR Pfam; PF01065; Adeno_hexon; 1.
DR ProDom; PD002815; Adeno_hexon; 1.
FT NON_TER 1
FT NON_TER 52
SQ SEQUENCE 52 AA; 5776 MW; CA787840339C61DC CRC64;

Query Match 75.3%; Score 143; DB 12; Length 52;
Best Local Similarity 83.9%; Pred. No. 7e-12;
Matches 26; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 GVINTETLTVKPKTGOENGWEKDATEFSDK 35
DB 3 GVINTETLTAKPKAGEENKWEKDATEFSDK 33

RESULT 5
Q80RI0
ID Q80RI0 PRELIMINARY; PRT; 952 AA.
AC Q80RI0;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hexon.
OS Human adenovirus type 5.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=28285;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KNH Ad00/5;
RA Na B.-K., Lee J.-Y., Kim J.-H., Kim W.-J., Kang C.;
RT "Sequence analysis of hexon protein of adenoviruses isolated in Korea.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF5421121; AAO24096.1; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000736; Adeno_hexon.
DR Pfam; PF01065; Adeno_hexon; 1.
DR ProDom; PD002815; Adeno_hexon; 1.
SQ SEQUENCE 952 AA; 107972 MW; 35440D6DC58E71C CRC64;

Query Match 75.3%; Score 143; DB 12; Length 952;
Best Local Similarity 83.9%; Pred. No. 1.7e-10;
Matches 26; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 GVINTETLTVKPKTGOENGWEKDATEFSDK 35
DB 419 GVINTETLTAKPKAGEENKWEKDATEFSDK 449
```

```
RESULT 6
Q93XC4 ID Q93XC4 PRELIMINARY; PRT; 226 AA.
AC Q93XC4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Thiol methyltransferase 2.
GN TMT2.
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3712;
RN [1]
RC SEQUENCE FROM N.A.
RP TISSUE=Leaf;
RA Attieh J., Djiana R., Etienne C., Sparace S.A., Saini H.S.;
RT "Cloning and functional expression of two plant thiol
RT methyltransferases: a new class of enzymes involved in the
RT biosynthesis of sulfur volatiles.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF387792; AAK69761.1; -
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000051; SAM_bind.
DR InterPro; IPR008854; TPMT.
DR Pfam; PF05724; TPMT; 1.
DR Methyltransferase; Transferase.
KW Methyltransferase; Transferase.
SQ SEQUENCE 226 AA; 25017 MW; 03389A8ED337F018 CRC64;

Query Match 33.7%; Score 64; DB 10; Length 226;
Best Local Similarity 48.0%; Pred. No. 1.6;
Matches 12; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 3 GKGVINTETLTQVKPKTQENGWEK 27
Db 13 GENIIPPEVAKPLPETVEGGWEK 37

RESULT 7
Q93V78 ID Q93V78 PRELIMINARY; PRT; 226 AA.
AC Q93V78;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Thiol methyltransferase 1.
GN TMT1.
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3712;
RN [1]
RC SEQUENCE FROM N.A.
RP TISSUE=Leaf;
RA Attieh J., Djiana R., Etienne C., Sparace S.A., Saini H.S.;
RT "Cloning and functional expression of two plant thiol
RT methyltransferases: a new class of enzymes involved in the
RT biosynthesis of sulfur volatiles.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF387793; AAK69762.1; -
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000051; SAM_bind.
DR InterPro; IPR008854; TPMT.
DR Pfam; PF05724; TPMT; 1.
DR Methyltransferase; Transferase.
KW Methyltransferase; Transferase.
SQ SEQUENCE 226 AA; 25127 MW; E6E339E518196798C CRC64;
```

```
Query Match 32.6%; Score 62; DB 10; Length 226;
Best Local Similarity 44.0%; Pred. No. 3.1;
Matches 11; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 3 GKGVINTETLTQVKPKTQENGWEK 27
Db 13 GENIIPPEVAKPLPETVEGGWEK 37

RESULT 8
Q80RH8 ID Q80RH8 PRELIMINARY; PRT; 944 AA.
AC Q80RH8;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hexon.
OS Human adenovirus type 3.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=45659;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=KNH Ad00/16;
RA Na B.-K., Lee J.-Y., Kim J.-H., Kim W.-J., Kang C.;
RT "Sequence analysis of hexon protein of adenoviruses isolated in
RT Korea.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF542125; AAO24100.1; -
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000736; Adeno_hexon.
DR Pfam; PF01085; Adeno_hexon; 1.
DR Pfam; PF03678; Adeno_hexon_C; 1.
DR ProDom; PD002815; Adeno_hexon; 1.
SQ SEQUENCE 944 AA; 106175 MW; 5DCFAF28BC56C964 CRC64;

Query Match 32.1%; Score 61; DB 12; Length 944;
Best Local Similarity 52.0%; Pred. No. 20;
Matches 13; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 5 GVINTETLTQVKPKTQENGWEKDA 29
Db 412 GIGFGHTFGQVKVKTDDANGWEKDA 436

RESULT 9
Q80RH6 ID Q80RH6 PRELIMINARY; PRT; 944 AA.
AC Q80RH6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hexon.
OS Human adenovirus type 3.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=45659;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=KNH Ad00/18;
RA Na B.-K., Lee J.-Y., Kim J.-H., Kim W.-J., Kang C.;
RT "Sequence analysis of hexon protein of adenoviruses isolated in
RT Korea.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF542127; AAO24102.1; -
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000736; Adeno_hexon.
DR Pfam; PF01085; Adeno_hexon; 1.
DR Pfam; PF03678; Adeno_hexon_C; 1.
DR ProDom; PD002815; Adeno_hexon; 1.
SQ SEQUENCE 944 AA; 106152 MW; 27CCB808BC56C965 CRC64;

Query Match 32.1%; Score 61; DB 12; Length 944;
```



DR GO: GO:0016787; F-hydrolase activity; IEA.  
DR GO: GO:0004525; F-ribonuclease III activity; IEA.  
DR GO: GO:0003723; F-RNA binding; IEA.  
DR GO: GO:0006396; F-RNA processing; IEA.  
DR InterPro: IPR001064; Crystallin.  
DR InterPro: IPR001410; DEAD.  
DR InterPro: IPR005034; DUF283.  
DR InterPro: IPR001650; Helicase\_C.  
DR InterPro: IPR003100; PAZ.  
DR InterPro: IPR000999; RNase III.  
DR InterPro: IPR007087; Znf C2H2.  
DR Pfam: PF033368; DUF283\_1; helicase\_C\_1.  
DR Pfam: PF02170; PAZ\_1.  
DR Pfam: PF00636; Ribonuclease\_3; 1.  
DR Pfam: PF00096; zf-C2H2; 1.  
DR SMART: SM00490; HELIC; 1.  
DR SMART: SM00535; RIBOC; 2.  
DR PROSITE: PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
DR PROSITE: PS00821; PAZ\_1.  
DR PROSITE: PS00517; RNASE\_3\_1; 1.  
DR PROSITE: PS00142; RNASE\_3\_2; 2.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
DR PROSITE: PS00157; ZINC\_FINGER\_C2H2\_2; 1.  
DR KW ATP-binding; Helicase; Hydrolase; Metal-binding; Zinc; Zinc-finger.  
SQ SEQUENCE 1627 AA; 180368 MW; DF7776452C2068F0 CRC64;

Query Match 30.5%; Score 58; DB 10; Length 1627;  
Best Local Similarity 35.5%; Pred. No. 93;  
Matches 11; Conservative 7; Mismatches 13; Indels 0; Gaps 0;  
QY 5 GVINTETLTQVKPKTQNGWEXDA 35  
DB 134 GLICASEATKICIERQEGKGLKEVVDATDQ 164

RESULT 14  
Q98EE6 PRELIMINARY; PRT; 265 AA.  
AC Q98EE6;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Amino acid ABC transporter, periplasmic amino acid-binding protein.  
DE MLL4278.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=NAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asanizu E., Kato T., Sasamoto S., Watanabe A., Ideesawa K., Ishikawa A., Kawashima K., Kimura T., Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti.";  
RL DNA Res. 7:331-338(2000).  
DR EMBL; AF003003; BAB50973.1; -;  
DR GO: GO:0030288; C:periplasmic space (sensu Gram-negative Bact. . .; IEA.  
DR GO: GO:0005215; P:transporter activity; IEA.  
DR GO: GO:0006810; P:transport; IEA.  
DR InterPro: IPR001311; SBP/glu\_receptor.  
DR InterPro: IPR001638; SBP\_bac\_3.  
DR Pfam: PF00497; SBP\_bac\_3\_1.  
DR SMART: SM00062; PEBB; 1.  
KW Complete proteome.  
SQ SEQUENCE 265 AA; 28814 MW; C8C0F8A74C27D359 CRC64;

Query Match 30.0%; Score 57; DB 16; Length 265;  
Best Local Similarity 46.4%; Pred. No. 17;  
Matches 13; Conservative 4; Mismatches 7; Indels 4; Gaps 1;  
QY 6 VINTET----LTQVKPKTQNGWEXDA 29  
DB 35 VVTENAYPPLQFIDPKTKGKIGNEYDA 62

RESULT 15  
Q91PL8 PRELIMINARY; PRT; 344 AA.  
AC Q91PL8;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hexon (Fragment).  
OS Human adenovirus type 3.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
OX NCBI\_TaxID=45659;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Ad3A724-G93;  
RA Mizuta K.;  
EL "Human Adenovirus Type 3 Hexon Partial Sequence.";  
DR EMBL; AB067674; BAB62520.1; -;  
DR GO: GO:0019028; C:Viral capsid; IEA.  
DR GO: GO:0005198; F:structural molecule activity; IEA.  
DR InterPro: IPR000736; Adeno\_hexon.  
DR Pfam: PF01065; Adeno\_hexon; 1.  
DR ProDom: PD002815; Adeno\_hexon; 1.  
DR NON\_TER 344 344  
FT NON\_TER 344 344  
SQ SEQUENCE 344 AA; 37995 MW; 74AF93E103DA58A6 CRC64;

Query Match 30.0%; Score 57; DB 12; Length 344;  
Best Local Similarity 66.7%; Pred. No. 23;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 15 VKPKTQNGWEXDA 29  
DB 291 IKVKTDDTNGWEXDA 305

Search completed: March 7, 2004, 13:37:40  
Job time : 39.0769 secs